AAMRL-SR-89-520

AD-A213 934

(2)

OPERATOR PERFORMANCE MEASURES FOR ASSESSING VOICE COMMUNICATION EFFECTIVENESS

K. L. DaPolito
M. E. Jones
S. B. Hottman

SYSTEMS RESEARCH LABORATORIES, INC.
A Division of Arvin/Calspan
2800 Indian Ripple Road
Dayton, Ohio 45440

July 1989

Approved for public release. Distribution is unlimited.

ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release. Distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
AAMRL-SR-89-500

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Systems Research Laboratories, Inc.

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Armstrong Aerospace Medical Research Laboratory (AAMRL/BBA)

6c. ADDRESS *(City, State, and ZIP Code)*
2800 Indian Ripple Road
Dayton, Ohio 45440-3696

7b. ADDRESS *(City, State, and ZIP Code)*
Wright-Patterson Air Force Base, Ohio 45433-6573

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F33615-85-C-0530

8c. ADDRESS *(City, State, and ZIP Code)*

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
62202F	7231	21	04

11. TITLE *(Include Security Classification)*
Operator Performance Measures for Assessing Voice Communication Effectiveness

12. PERSONAL AUTHOR(S)
K. L. DaPolito, M. E. Jones, S. B. Hottman

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 10/87 TO 9/88

14. DATE OF REPORT *(Year, Month, Day)*
1989 July

15. PAGE COUNT
212

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	05	
17	02	.1

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Voice Communication Effectiveness
Operator Performance
Speech Intelligibility Information Processing

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report documents the design and rationale of a voice communication effectiveness measure, for the assessment of voice communication systems where information is passed two ways over a communication link. New response tasks other than word intelligibility were required. The communication effectiveness measure selected and integrated into a research scenario is based upon classic information theory. The scenario can be used alone or in conjunction with secondary tasks. A secondary task was selected for use with the communication scenario, and the integration of the primary and secondary tasks for voice communication effectiveness is documented.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Thomas Moore, Ph.D.

22b. TELEPHONE *(Include Area Code)*
(513) 255-3607

22c. OFFICE SYMBOL
AAMRL/BBA

SUMMARY

This report documents the selection of a voice communication effectiveness measure for implementation into the Armstrong Aerospace Medical Research Laboratory, Biological Acoustics Branch (AAMRL/BBA), Performance and Communications Research and Technology (PACRAT) facility. Research to date by AAMRL/BBA has been concerned with the intelligibility of individual words transmitted and received in the presence of various noise and interfering modulations. The research has proceeded to the point where data are needed on the effects of interfering with interactive voice communications where relevant information is passed both ways over a communication link. To gather these data, new response tasks were required other than word intelligibility tasks presently used.

Based on a literature review of speech intelligibility, human performance/workload, and information theory, a performance task was selected for the measure of voice communication effectiveness. The selected performance task is an interactive voice communication scenario with high verbal demands. The communication scenario (primary task) utilizes a database of confusable words. The words in the database have been analyzed for mutual information and entropy. Researchers can employ the scenario as an independent voice communication effectiveness measure, or it can be used in conjunction with a secondary task. The secondary task, selected for use with the primary task, is a compensatory tracking task. Dependent variables from the primary and secondary task include response time, errors, number of requested repeats, timeouts, and root-mean-square error.

As the voice communication effectiveness measure is implemented on the PACRAT facility, research will be required to determine its validity and reliability. Additional recommendations concern the expansion of the family of secondary tasks.

PREFACE

This work was conducted by personnel of Systems Research Laboratories, Inc. (SRL), under Contract No. F33615-85-C-0530 with the Harry G. Armstrong Aerospace Medical Research Laboratory (AAMRL), Human Systems Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio. The work activities summarized in the report were performed under Program 62202F, Aerospace Biotechnology, Project 7231, Biodynamics of Air Force Operations.

The authors wish to acknowledge the important contributions to this report made by several people. Dr. Thomas Moore and Mr. Richard McKinley of AAMRL/BBA provided substantial direction and contributions to the scenario. From SRL, several individuals contributed to this report. Mr. Ronald Dallman compiled the database of pilot communications for the scenario, and Ms. Terese DeSimio generated the scenario sentences and analyzed them for entropy and mutual information. Dr. William Perez and Mr. John Simons provided guidance on the primary and secondary task applications. Mr. Donald Green, a former U.S. Air Force test pilot, provided an operational perspective which was important for the scenario development. The authors also wish to thank Mr. Alva Karl for his continuing encouragement and support during the performance of this task.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

Section 1
INTRODUCTION

1.1. BACKGROUND

Air and ground crew voice communications may be degraded by a variety of environmental and systems factors. Such factors may include electrical or acoustical noise, radio interference, jamming, communication signal processing, as well as various other forms of interference. As communication is a vital part of the flight environment, research activities attempting to identify and quantify potentially degrading elements of the operational environment must be maintained. Analytical studies of communication system performance and the effect of environmental influences on those systems are necessary. Such studies are possible in controlled laboratory environments where special instrumentation can be used to create the elements of the human factors and communication system networks being investigated.

The Harry G. Armstrong Aerospace Medical Research Laboratory/Biological Acoustics Branch (AAMRL/BBA) has been engaged in a long-term research program sponsored by the Air Force Electronic Warfare Center (AFEWC) and the Joint Electronic Warfare Center (JEWC) to conduct such investigations. The majority of this research has investigated the effects of interference on the intelligibility of individual words transmitted or received over a communication channel. Data collection instruments have included the Modified Rhyme Test (MRT), the Diagnostic Rhyme Test (DRT), and the Coordinate Response Measure (CRM).

1.2. PURPOSE OF RESEARCH

The results of AAMRL/BBA's research indicate a need for new research on the effects of interference on interactive voice communication situations--in other words, situations where information is passed two ways over a communication channel. Specifically, the amount of intelligibility required to perform various tasks and the amount of time required to perform those tasks must be determined. To investigate this area, new response tasks other than the standard intelligibility tests previously mentioned are needed. AAMRL/BBA requires that these new metrics employ standard measures of performance such as total time needed to complete a task, number of repeats requested, and number of errors made in completing a task.

To develop these new performance metrics, Systems Research Laboratories, Inc. (SRL), was tasked with the following: (1) a survey of the literature concerning performance measures of voice communication effectiveness (VCE), (2) an evaluation of various performance measures of VCE, and (3) the development of a new VCE metric to be implemented in AAMRL/BBA's test facility. This report describes SRL's efforts associated with these tasks. Sections 2 through 6 describe the results of the literature survey. Section 7 describes the performance measures considered for implementation in the VCE test facility. Conclusions and recommendations for future research are provided in Section 8.

1.3. VCE TEST FACILITY

The Performance and Communications Research and Technology (PACRAT) laboratory, located in the basement of Building 441 at Wright-Patterson Air Force Base (WPAFB), is the facility used by AAMRL for VCE research.

The PACRAT facility currently consists of seven individual subject communication stations, one experimental control station, and a high intensity sound system capable of duplicating operational acoustical environments. The seven subject stations are housed in a single large reverberation chamber. The control station and the sound system control panel are located in a room adjacent to the reverberation chamber.

Each of the seven subject stations is a modified aircraft simulator shell with communication, display, input, and data acquisition capabilities as depicted in Figure 1. The display devices include four CRT screens: a large 13-inch screen, and three smaller 9-inch screens (Figure 2). All CRT screens have color and graphics capabilities. Each subject station has two communication addresses to which it will respond. One address is common to all stations. By using this address and a single message, all stations can simultaneously receive the same information. The second address system is specific to individual stations. Using this address, different messages may be simultaneously sent to different stations. The subject stations also have two different response systems. The first system consists of 60 pushbuttons, 20 per small CRT (five on each side), as can be seen in Figures 1 and 2. The second response system is an F-16 style force joystick with push-to-talk and electric trim switches. This system may be used to respond to information displayed on the large CRT screen. Each of the seven subject stations is also compatible with standard Air Force headgear and respiration systems.

The experimental control station or central processing unit controls each of the individual stations and conducts the individual testing sessions. The control unit is responsible for presenting test material, monitoring participant (both sender and receiver) activity, and recording, storing, and analyzing subject responses.

Figure 1. Photo of a Subject Test Station in the VCE Test Facility

Figure 2. Photo of the Display Devices on the Subject Test Station (The large 13-inch CRT screen and the three smaller 9-inch screens all have color and graphic capabilities.)

The PACRAT sound system is comprised of a noise generator and a spectrum shaper capable of generating almost any desired noise environment within the human audio frequency range. This permits accurate reproduction within the PACRAT test chamber of the ambient and environmental noise conditions of specific operational situations. Speaker banks are located in the specially designed and constructed reverberation chamber. This chamber is constructed to maximize the uniformity of the level of noise distributed throughout the room.

1.4. LITERATURE SEARCH

SRL utilized existing government and commercial databases to provide a survey of the literature pertaining to the evaluation of voice communication effectiveness. Based upon the research requirements of AAMRL/BBA and test facility characteristics, several topic categories were selected for search. These categories included: speech communication, human information processing, operator performance measures, tactical scenarios, and communication theory. Table 1 displays the subcategories searched within each topic area. Computer searches of each area were conducted on DIALOG's Aerospace and Conference Proceedings Index databases; the NASA and NTIS technical reports databases; and NERAC's Engineering Index, Biological Abstracts, and Inspect databases on conference papers, journals, and news reports. Manual searches of citation and reference indices included the International Technical and Scientific Index, Medicus Index, Psychological Abstracts Index, Science Citation Index, and Social Science Citation Index. Manual searches of the holdings of the Wright Research Development Center (WRDC) Technical Library, the Wright State University library, and the University of Dayton library were also conducted. The literature search yielded several hundred sources which are documented in the bibliography attached to this report. Sections 2 through 6 describe the results of the literature search in detail.

1.5. VCE PERFORMANCE MEASURE DEVELOPMENT

Based upon the results of the literature review and the constraints/requirements of the PACRAT test facility, three potential performance measures for VCE research were developed. Two of these tasks have been selected for implementation. Section 7 gives a detailed description of each of these tasks.

TABLE 1. LITERATURE SEARCH SUBCATEGORIES

Speech Communication

- Verbal Communication
- Noise and Speech
- Speech Intelligibility Measures
- Communication Research
- Applied Aviation Noise and Communication Research
- Synthetic Speech Technology

Human Information Processing

- Perception
- Memory
- Learning
- Attention
- Language Specialization
- Decision Making
- Problem Solving
- Auditory Information Processing
- Models of Processing

Operator Performance Measures

- Performance Measures of Behavior
- Performance Measures of Psychological/Psychophysiological Processes
- Operator Workload
- Existing Performance Batteries

Tactical Scenarios

- General Tactical Profiles
- Communications Oriented Profiles

Communication Theory

- Language
- Speech Analysis
- Information
- Communication Logic
- Mathematical Theories of Communication

Section 2
SPEECH COMMUNICATION

2.1. INTRODUCTION

Available speech communication literature describes a variety of topics. These topics include the description of the physical components of speech--for example, frequency and intensity (Chapanis, 1965; Kryter, 1984, 1985); the linguistics of speech (i.e., phonemes, syllables, words, vocabularies, and messages); the methods of speech production and articulation; auditory perception and information processing (Carterette and Friedman, 1976; Cole, 1980; Hawkins and Presson, 1986; Jusczyk, 1986; McCormick, 1976); and the effectiveness of speech communication (Chapanis, 1965; Harris, 1979; Kryter, 1984, 1985; Van Cott and Kinkade, 1972; von Gierke and Nixon, 1985).

Since this report is concerned specifically with the effectiveness of verbal speech communication, the following sections focus on research documenting that topic: Section 2.2 defines human verbal communications, Section 2.3 describes the various methods of measuring speech intelligibility, Section 2.4 discusses research on interactive voice communication, Section 2.5 describes synthetic speech, and Section 2.6 reviews the literature on aviation noise and communication research. It should be noted that there exists very little published research on interactive communication. However, many components of interactive communication such as speech intelligibility, environmental effects, and operational factors are well documented.

2.2. VERBAL COMMUNICATION

Human verbal communication exists in two forms: unidirectional (i.e., noninteractive) and interactive communications. Unidirectional communication describes communication in which the person to whom the message is addressed is a passive recipient. The receiver of this form of communication can in no way affect the communicator, the communication process, or the content of the message that is received. Examples of unidirectional communication include speeches, lectures, and television broadcasts. Interactive communication describes situations in which more than one of the participants are both senders and receivers of information. Interactive communication is not passive. Participants can affect the other communicators, the process itself, and the content of the message. Examples of interactive communication include two-way radio transmissions, arguments, telephone conversations, and human-computer dialogue.

Verbal communication, both unidirectional and interactive, is crucial to the successful performance of innumerable tasks in most aerospace operations. Measurement techniques for determining the adequacy of verbal communication then become important. Measurement techniques described in the literature include both "voice communications effectiveness" and "speech intelligibility" measures. These two terms are often used interchangeably. "Voice communications effectiveness" can be defined as the efficacy of verbal communication while "speech intelligibility" may be defined as the understanding of spoken words. Despite the similarity in these definitions, there is an important difference in their meaning. The term "voice communications effectiveness" not only includes the intelligibility or understandability of speech, but also implies that there is a response (or some performance) made by the receiver based upon the intelligibility of the message.

Both the intelligibility of speech and the effectiveness of voice communications can be influenced by a variety of environmental, human, message, and system factors. Environmental factors include noise, vibration, acceleration, stressors, and task requirements. Human influences include speech habits, dialects, word usage, language familiarity, hearing loss, communication experience, motivation, workload, and emotional state. Elements of the message which influence voice communications effectiveness include vocabulary size, vocabulary familiarity and frequency, message redundancy, message presentation, and context. Equipment factors include interference with the clarity, volume, etc., of the speech signal, and interference with the subject's auditory capabilities.

2.3. SPEECH INTELLIGIBILITY MEASURES

A variety of standardized methodologies are described in the literature for measuring the performance of voice communication systems. Methodologies exist for measuring both the entire communication system and its various individual elements. Both subjective measures, in which the percentage of a given speech sample that is correctly perceived by a receiver, and physical measures of the system and the environment exist. Relating the subjective measures to the physical measures allows the effectiveness of speech communication to be assessed.

2.3.1. Physical Predictors of Intelligibility

Physical measures of the system and environment used to predict speech intelligibility include the A-weighted sound level [dB(A)], speech interference level (SIL), noise criteria (NC), and the articulation index (AI). SIL, dB(A), and NC measures will not be discussed here as they do not provide very comprehensive assessments of intelligibility (see Webster, 1979 for descriptions of these measures).

2.3.1.1. The Articulation Index

The AI is perhaps the most widely used of the physical predictors of speech intelligibility. Calculation of the AI is based upon determination of a weighted signal-to-noise ratio from the level of the speech signal and the noise in the environment. The difference between the level of the speech and the level of the noise is measured in 20 contiguous bands of frequencies. These frequency bands contribute equally to speech intelligibility when all are at optimal gain. The average difference between signal and noise (across all bands) is then normalized to yield a value between 0 and 1.0. A value of 0 indicates the listener will rarely be able to understand speech in the given environment, while a value of 1.0 indicates potentially perfect perception by the listener. The American National Standard Methods for the Calculation of the Articulation Index (ANSI S3.5-1969) describes detailed instructions for calculation of the AI.

2.3.1.2. The Speech Transmission Index

Steeneken and Houtgast (1980, 1981; Houtgast and Steeneken, 1981) have developed a Speech-Transmission Index (STI) which is an extension of the AI. The STI is based on the Modulation Transfer Function (MTF) of a transmission channel, and is used as a physical method for measuring speech-transmission quality. A study by Steeneken (1987) compared the STI with the DRT and a Consonant-Vowel-Consonant (CVC) word tests to obtain speech intelligibility scores for diagnostic information related to the type of deterioration to which the speech signal was subjected. He found that the STI provided better diagnostic information for the evaluation and classification of speech channels than either the DRT or CVC word tests.

2.3.2. Subjective Measures of Intelligibility

Measures of intelligibility which are based upon psychoacoustic measurements of the communication system or the environment can be divided into four main classes: (a) nonsense syllables, (b) spondaic words, (c) sentences, and (d) monosyllabic words (Chapanis, 1959). Tests using monosyllabic words are further subdivided into phonetically balanced (PB) word tests (ANSI, 1960) and rhyme tests like the Fairbanks Rhyme Test (Fairbanks, 1958), the Modified Rhyme Test (MRT) (House, Williams, Hecker, and Kryter, 1965), and the Diagnostic Rhyme Test (DRT) (Voiers, 1968, 1977). The DRT has since been computerized by the U.S. Army to test armored vehicle intercommunications systems (Mayer, 1985).

2.3.2.1. Nonsense Syllables

Nonsense syllables (e.g., monz, nan, fook) have been successfully used to determine the effectiveness of specific communication devices in transmitting particular speech sounds (Beranek, 1949; Chapanis, 1959). Figure 3 depicts the relativity of nonsense syllables to words and sentences for evaluating such transmission equipment. Unfortunately, using nonsense syllables requires extensive training for both talkers and listeners. Talkers have to learn to correctly pronounce the fundamental speech sounds that comprise the test, and listeners must learn to recognize these sounds and be able to record the associated phonetic symbols.

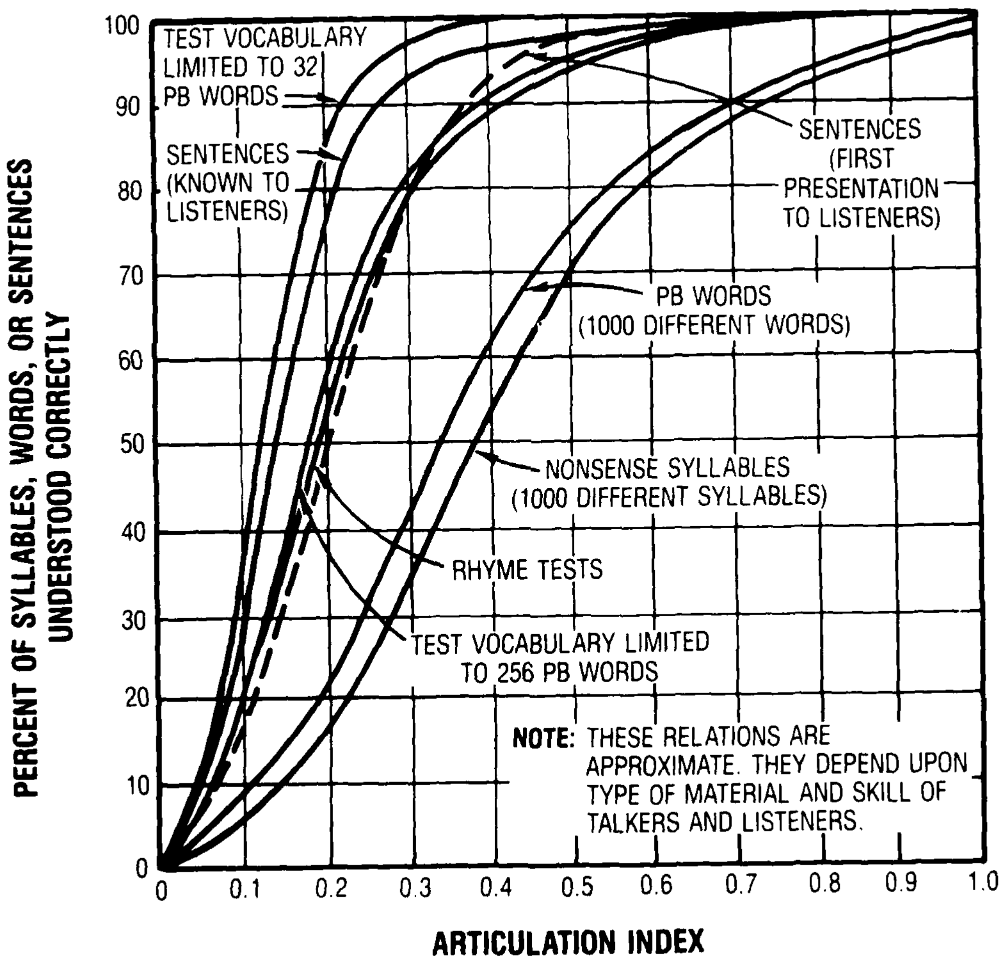

Figure 3. A Comparison of Speech Intelligibility Measures with the Articulation Index (ANSI S3.5-1969)

2.3.2.2. Spondiac Words

Spondiac words, or spondees, are primarily used to determine speech level settings on equipment to achieve the threshold of detection by listeners. The spondees themselves are two-syllable words which are spoken with equal stress on each syllable (e.g., airplane, woodchuck). These words reach the listener's threshold of hearing within a very narrow intensity range. This allows for high precision in the experimenter's measurements.

2.3.2.3. Sentences

Sentence tests, like spondiac word tests, are of rather limited use for testing communication equipment. Because sentences have certain inherent characteristics (meaning, context, rhythm) which words and syllables do not have, sentence tests typically yield very high intelligibility scores. Since these scores are usually high, communication systems must differ greatly to achieve a significant difference in scores (Beranek, 1949). Sentence tests are, however, useful for testing the maintenance of loudness levels, and for evaluating the rate, inflection, and stress patterns of talkers' speech. The sentence lists used for testing are normally one of two kinds. They are either questions requiring an answer from the listener (e.g., What letter comes after "Q"?), or statements which must be recorded by the listener (e.g., Take the cards from the deck, you bum). For questions, wrong answers are scored as errors. For statements, only five key words (predetermined and underlined for the talker) are checked for correctness. Cliches, proverbs, popular phrases, and very frequently used words are not used in the sentences (Beranek, 1949; Chapanis, 1959; Egan, 1948; Kalikow and Stevens, 1977; Kryter, 1972).

2.3.2.4. Monosyllabic Words

The most commonly used test materials for determining speech intelligibility are monosyllabic or one-syllable words. As stated previously, monosyllabic words are used in the PB word tests and rhyme tests.

2.3.2.4.1. Phonetically Balanced Words

Typical PB word lists consist of 50 words each. A set of 20 of these lists are provided by the USA Standard Method for Measurement of Monosyllabic Word Intelligiblity (ANSI S3.-1960). The frequency of occurrence of the types of speech sounds (e.g., fricatives, glides, nasals) are approximately the same as in normal everyday speech; hence, they are deemed "phonetically balanced." The words in each list are also approximately equal in difficulty; thus, if an average intelligibility score of 50 percent is achieved by a test group, then very few of the words will be extremely easy or difficult to understand (Beranek, 1949). The PB word lists of the ANSI S3.2-1960 should each be randomly reordered before each use. For best results in testing, all 20 lists (1000 different words) should be used (see Figure 3).

2.3.2.4.2. Modified Rhyme Test

The MRT is the most commonly used monosyllabic word test. The MRT normally consists of 50 numbered sets of six words on an answer sheet for the listener, and 50 numbered single words, one word taken from each set of the six words on the listener's list, for the talker. The talker announces the key word in a carrier sentence like "Number (of the key word), you will mark the (key word) please" (Kryter, 1972), or a carrier phrase like "Number (of the key word) is (key word)" (House et al., 1965). Of the set of 50 words, 25 sets are such that the final consonantal element is varied, and 25 sets vary the initial consonantal element. An example of each type of set is as follows:

1.	bat	bad	back	bass	ban	bath
2.	look	took	shook	cook	hook	book

The listener's answer sheet is scored by counting the number of words correctly marked for the test. This amount is then corrected for chance guessing by using the following formula (Kryter, 1972):

$$\% \text{ correct} = [\text{No. right} - (\text{No. wrong}/5) \times 2]$$

Like the PB word lists, each MRT list should be randomly generated between sets and within each set for each test. The MRT words are not phonetically balanced to reflect everyday usage, but the MRT is still considered useful and efficient. This is because it requires perception of consonantal sounds. These sounds are difficult to transmit successfully and, therefore, important to intelligibility (Kryter, 1972).

2.3.2.4.3. Diagnostic Rhyme Test

The DRT word lists were developed to test consonant discriminability with six features of a phonemic taxonomy: (1) voicing, (2) nasality, (3) sustention, (4) sibilation, (5) graveness, and (6) compactness (Voiers, 1968, 1977). There are 96 rhyming word pairs used in the DRT that differ phonemically on their initial consonants. An example of the word pairs for each feature are as follows (Voiers, 1977):

Feature	Example A	Example B
Voicing	Dint-Tint	Zoo-Sue
Nasality	Nip-Dip	Moot-Boot
Sustention	Thick-Tick	Foo-Pooh
Sibilation	Sing-Thing	Juice-Goose
Graveness	Fin-Thin	Moon-Noon
Compactness	Gill-Dill	Coop-Poop

Together the six phonemic perceptual features provide an overall gross measure of speech intelligibility; although, if necessary, they can be measured separately. The word pairs are usually presented so that each feature appears twice to each listener for each trial. The listener is given a pencil and a list of word pairs to be announced, and then marks out the one word of each pair that he/she perceives to have been spoken. The overall speech intelligibility score is adjusted for guessing with the follow correction formula (Voiers, 1977):

$$S = \frac{100\,(R - W)}{T}$$

where S is the adjusted percent of correct answers, R is the number of right answers marked, W is the number of wrong answers marked, and T is the total possible number of correct answers.

2.4. INTERACTIVE VOICE COMMUNICATIONS RESEARCH

The major focus of research on interactive voice communications has been on human-to-computer interaction in voice interactive systems. The small amount of work that has been done on human-to-human interactive voice communications has been in conjunction with basic research on human-computer interactions. Chapanis (1971) points out that before a truly interactive computer system can be developed, it is necessary to better understand the interaction between human beings engaged in communication.

Chapanis, Ochsman, Parrish, and Weeks (1972) described experiments to study interactive communication of two-person teams during cooperative problem-solving. They studied these effects using four communication modes: (a) typewriting, (b) handwriting, (c) voice, and (d) "communication-rich." The "communication-rich" mode entailed two subjects communicating, face-to-face, in any way they wanted. In the other modes, the subjects were separated by partitions which included holes to pass notes written during tests of the handwriting mode of communication. The typewriting mode (using teletypewriters) was further split into two groups with one group composed of inexperienced typists and the other group composed of experienced typists (i.e., typists having completed at least a 1-year course of high school typing).

For each two-person team, one subject was designated as the information source or "source," and the other subject was deemed the information seeker or "seeker." The source subject was to be considered as a hypothetically ideal computer, and the seeker subject as the user of that computer. Two problems were used for the tests: (1) an equipment assembly problem, and (2) a geographic orientation problem. In the equipment assembly problem, the seeker was to assemble a trash can carrier. In the geographic orientation problem, the seeker was to find the office or residence address of a physician closest to a hypothetical home address. Three dependent variables were measured: (a) time to arrive at a solution, (b) behavioral measures of activity, and (c) linguistic measures.

The results indicated that the two voice communication modes were significantly superior for interactive communication. Overall, subjects in the "communication-rich" mode condition took the shortest time to arrive at a solution, with a mean time of just less than 30 minutes. Subjects in the voice mode condition took just under 35 minutes to arrive at a solution (this was not significantly different from the communication-rich mode). Subjects using the other modes took nearly twice as long to solve their problems. The handwriting mode was superior to the typewriting mode, and experienced typists were slightly faster problem solvers than inexperienced typists.

Results of the behavioral measures suggest that the source and seeker subjects using the voice-only mode spent almost equal amounts of their time searching for and sending data to solve the problem; however, the source subjects did spend more time in sending data for the equipment assembly problem.

The linguistic measures (Chapanis, Parrish, Ochsman, and Weeks, 1977) showed that many more messages, sentences, and words were used in the voice and "communication-rich" conditions than in the other conditions. On the average, the subjects using the oral modes talked about 183 words per minute; however, the source subjects used longer messages (11.8 words per message versus 7.9 words) and longer sentences (6.7 words per sentence versus 5.0 words) than did the seeker subjects.

Chapanis and Overbey (1974) did another similar experiment using 32 college students for subjects. The subjects were again assigned jobs as either seekers or sources in one of the two adjacent rooms used during the previous experiment. The subjects used a speaker-microphone system and/ or a teletypewriter system for communication. There were four different configurations of communication modes used during the tests. In a seeker-source relation, the four modes were:

(a) voice-voice (V-V), (b) voice-typewriter (V-T), (c) typewriter-voice (T-V), and (d) typewriter-typewriter (T-T).

Four problems were used in this experiment. Two were the same as in the previously mentioned experiment. The other two problems consisted of either an information retrieval problem or an object identification problem. In the information retrieval task, the seeker was to find five citations of different newspaper articles relevant to a given topic from a portfolio of newspapers given to the source. In the object identification task, the seeker was to identify and obtain a replacement for a small pilot light socket from a large number of different sockets kept by the source.

The results of these experiments confirmed the earlier findings that a voice mode of communication was significantly better for problem solving than a typewriting mode. In the seeker-source relationship, the rank order of the communication modes were as follows: (1) V-V, (2) V-T, (3) T-V, and (4) T-T. The average message lengths used for solving the problems were about five times faster (3.0 messages per minute) for the V-V mode than for the T-T mode (0.6 messages per minute).

Chapanis (1975, 1976) again tested interactive communication modes. In this experiment, ten different communication modes (five with voice and five without voice communications) were tested. Two generalizations resulted from this experiment: (1) that communications problems are solved significantly faster when verbal communication is allowed, and (2) that problems are solved equally well in voice-only and face-to-face modes.

2.5. SYNTHETIC SPEECH AND INTELLIGIBILITY

A number of experiments have investigated the effects of synthetic speech systems on intelligibility. Porubcansky (1985) describes the increasing interest in automated speech technology for use in Air Force aircraft. Of the two types of speech synthesis production (i.e., phonemic synthesizer and encoded speech synthesis system), the encoded speech synthesis systems produce the most intelligible speech. Thus, the Air Force has focused its research programs on the development of the best possible speech synthesis system using speech waveform encoding techniques, such as linear predictive coding (LPC).

LPC and other waveform encoding techniques have been investigated for voice communication effectiveness by McKinley and Moore (1986). McKinley and Moore measured the speech intelligibility in simulated aircraft cockpit noise with ten subjects by using the MRT. They found that different audio bandwidths and bit error rates significantly effected the speech intelligibility for the encoding techniques used.

The DRT is still "widely used to evaluate digital voice systems" (Schmidt-Nielsen, 1987). The problem that many researchers have with using the DRT is that there is no reference frame for interpreting DRT scores for every day or operational performance measures. Schmidt-Nielsen compared DRT scores with intelligibility levels of the International Civil Aviation Organization (ICAO) spelling alphabet (e.g., Alpha, Bravo, Charlie, etc.) using an LPC algorithm. Her results showed that ICAO intelligibility remained high until the DRT scores fell below 75 percent, at which point the ICAO intelligibility dropped off quickly until the DRT scores reached 50 percent. At the 50 percent DRT level, the ICAO intelligibility level was about half.

Slowiaczek and Nusbaum (1985) examined the effects of speech rate and pitch contour on the perception of speech. The results indicated that speech rate influenced intelligibility more than pitch contour. Greene, Logan, and Pisoni (1986) used the MRT to evaluate intelligibility of eight off-the-shelf text-to-speech systems as compared to natural speech. The results showed only one of the systems, DECtalk-Paul, was comparable to natural speech.

2.6. AVIATION NOISE AND COMMUNICATIONS RESEARCH

2.6.1. General Aviation and Communications Noise Research

The various studies utilizing aircraft noise in conjunction with issues of speech communication fall into two groups--studies done with a point of reference from outside an aircraft (Arnoult and Voorhees, 1980; Fröhlich, 1981; Kryter and Williams, 1965; Pollack, 1958; Webster, 1965; Williams, Mosko, and Greene, 1976), and those done from inside the aircraft/cockpit (Arnoult, Voorhees, and Gilfillan, 1986; Lacey, 1973; Pratt, 1981; Wheeler and Halliday, 1981; Williams, Forstall, and Greene 1971). The studies of interest in this report are those done with a point of reference inside the aircraft.

Williams, Forstall, and Greene (1971) used an in-flight manikin to evaluate the communications effectiveness of three different helmets. Speech intelligibility tests (viz., MRT) were transmitted to six subjects along with the manikin in an airborne C-45 aircraft. Later, on the ground, the flight was simulated by reproducing the aircraft cabin noise in the laboratory. The same six subjects and two groups of ten listeners were retested by replaying the recordings made with the manikin. The results showed very little difference in scores for the two test situations. Williams et al. concluded that in-flight manikins could indeed be used to test the communication effectiveness of flight helmets. The data also seem to suggest that intelligibility tests can be performed via laboratory simulations with good results, and at less expense than actual in-flight measures.

To evaluate the effectiveness of a voice display mode for jet aircraft in an air combat maneuver environment, Lacey (1973) used speech intelligibility tests (i.e., MRT and operational word lists) in conjunction with both simulated aircraft noise and background speech interference. Results showed that the pilots who participated as subjects understood approximately 65 percent of the MRT words and 89 percent of the operational words. Based upon these results, Lacey suggests that a voice advisory system may be feasible during air combat maneuvers.

Wheeler and Halliday (1981) describe a laboratory evaluation of an active noise reduction (ANR) system for flight helmets. Subjects' performance on a speech intelligibility test (i.e., an MRT) was recorded in various conditions of background aircraft noise. One half of the subjects performed the intelligibility test while wearing the ANR helmet. Depending on the specific noise condition, the ANR system appeared to reduce noise 15 to 20 dB(A).

Pratt (1981) used simulated aircraft noise (viz., helicopter) and tested subjects using the MRT and the Clarke's Vowel Test (CVT) to measure the effectiveness of an automated multiple choice intelligibility testing system. The CVT uses single syllable words (where only the vowels change between words). As in the MRT, subjects had to choose each keyword from a group of words, but only five words instead of six. The results showed that there was no significant difference between the automatic and the manual tests. The CVT scores did not fare so well, whereas there was a significant difference between the manual and automatic tests.

Arnoult, Voorhees, and Gilfillan (1986) investigated the effects of annoyance on speech intelligibility in various backgrounds of simulated helicopter cabin noise. The test materials used were complete sentences. The sentence tests were developed following the recommendations of Hudgins, Hawkins, Karlin, and Stevens (as cited in Arnoult et al., 1986), with the exception that all sentences were to be answered as either true or false. Altogether, 160 sentences were made. These were presented in groups of ten for 16 sets, each set having five true and five false statements, and randomly arranged in each set. The sentences were prerecorded by a male speaker. The simulated helicopter cabin noise was composed of two components: (1) a pink noise (PN) broadband signal, and (2) one of three pure tones (PT) at 650, 1900, or 5000 Hz. These components were then generated, in all combinations, at four sound levels [i.e., 0, 60, 70, and 80 dB(A)]. The sentences were presented at either 50 or 55 dB(A). The results indicated that both noise sources, PN and PT, and their interactions were significant ($p < .001$) regarding intelligibility and annoyance. The PN component had relatively more effect on intelligibility loss, and the PT components caused more annoyance.

Interactive communication degradation from audio jamming is an important concern for the aerospace community. A classic test called the Michigan Map Test was developed in the 1950s (cited in Bennighof, Farris, Lauderdale, Richard, and Wild, 1978). The map is made up of a criss-cross pattern producing a field of diamond shapes with each corner (representing a town) designated by a letter from a phonetic alphabet. A talker attempts to guide a listener through the diamond grid from one of 972 predesignated possible routes available for use. Each route represents 9.925 bits of information. The route is transmitted when the jamming begins and the time is measured for the receiver to travel to six towns. The measured time is the jamming effectiveness measure for each jammer/signal ratio. The guideline of performance goes from a base reference point of 2 seconds with no jamming to 20 seconds (maximum) with jamming.

2.6.2. AAMRL/BBA Research

The Communication Evaluation Facility, now known as the Voice Communication Research and Evaluation System (VOCRES), located in the Biodynamics and Bioengineering Division of AAMRL (McKinley, 1980, 1981), has been used extensively for testing the effectiveness of communication equipment in various noise and jamming environments. Using this facility, Moore, McKinley, Mortimer, and Nixon (1978) evaluated the word intelligibility of two modulator/demodulator (modem) systems of a spread spectrum communication system in the presence of simulated F-15A cockpit noise. Moore et al. used various jamming conditions with cockpit noise while administering the MRT. The results showed that increased jamming with cockpit noise did degrade the MRT scores, and that the advantages of either modem were case specific depending on the jammer-to-signal power ratio.

Additional jamming research has been conducted in the VOCRES facility. Moore (1981) examined the comparative effectiveness of five different types of jammers. Two types of tests materials were used to measure intelligibility: (1) the MRT, and (2) a more operationally realistic word test developed by Ascher et al. (cited in Moore, 1981). The results of both intelligibility tests showed that the dual FM swept tones jamming signal was the most effective. Also, during this study, reported later by Nixon, McKinley, and Moore (1982), listeners were evaluated for training effects on increased intelligibility of jammed words. The results indicated that training did improve the listeners' ability to recognize jammed words.

Research on the effect of aviation cockpit noise on word intelligibility without jamming has included the evaluation of various radio systems (Moore, McKinley, and Mortimer, 1979), and in-flight headsets (Prohaska and Nixon, 1984). The ARC-34 and ARC-164 transceiver radio systems were tested using the MRT and three levels (i.e., 95, 105, and 115 dB) of cockpit noise in the VOCRES. The ARC-164 radio was found to perform better in all three noise levels. Various nonstandard in-flight headsets were compared against each other and the standard in-flight headset (viz., H-157), again by obtaining MRT scores in one-third octave bands of noise. Results favored the nonstandard headsets, but also suggested that the standard headset provided more attenuation, especially at the higher frequencies tested.

In response to reports from aircrew members that positive pressure breathing affects voice communication, Nixon (1984) studied positive pressure breathing under various conditions of simulated aircraft noise in the VOCRES. Using the MRT, Nixon found that speech intelligibility was not significantly degraded until the simulated cockpit noise reached 115 dB. There was, however, a trend across the other noise conditions suggesting an inverse relation between breathing pressure and intelligibility.

In addition to these studies, Moore and McKinley (1986) presented a review that described a number of speech related studies being conducted by AAMRL/BBA. In their review, and pertinent to the last study on pressure breathing, is a brief discussion concerning the effects of acoustic-phonetic changes from acceleration. This is relevant because, as Nixon (1984) related, communication in an actual aircraft is done under varying degrees of G-force along with the pressure breathing experienced by aircrew members. Moore and McKinley also presented other AAMRL/BBA data related to modern issues of speech coding, and mentioned some of the experiments being done on synthetic speech and speech recognition devices.

Section 3
HUMAN INFORMATION PROCESSING

The information processing requirements of piloting tasks are great. Modern aircraft, particularly jet fighters, allow pilots access to a great amount of information, yet often give them little time to process it. This information may be displayed and acted upon in a number of ways. In aircraft, visual and auditory displays of information are most common. These displays may require spatial or manual transformations of information and manual and/or vocal responses. The result is the requirement to perform complex, difficult, highly cognitive tasks in limited amounts of time. Although a thorough review of the human information processing system is beyond the scope of this report (see Boff, Kaufman, and Thomas, 1986; Lindsay and Norman, 1977), the following section discusses some of the more critical aspects of the human information processing system as it relates to the performance and measurement of pilot communication tasks. Section 3.1 briefly discusses research on auditory information processing, and Section 3.2 describes the major processing models upon which operator performance and workload assessment theories are based. The development of a performance metric assessing voice communication effectiveness should be based upon the findings described in the literature.

3.1. AUDITORY INFORMATION PROCESSING

A majority of a pilot's communication activities involve some auditory component. Many tasks require the pilot to identify and respond to incoming verbal messages and comments. Unfortunately, a large amount of incoming verbal information may be degraded due to electrical and acoustical noise, radio interference, and jamming (McKinley, 1980). Such environmental interference often increases the difficulty of the communication task, perhaps even making successful completion of the task impossible. The remainder of this subsection discusses the ability of humans to process complex auditory information, specifically speech sounds. Topics to be covered include auditory attention and auditory memory.

3.1.1. Auditory Attention

Research on auditory attention was initiated in the early 1950s based on the need to better understand the communication behavior of air traffic controllers and pilots who needed to respond quickly and accurately to a wide range of both visual and auditory information. A majority of this research has focused on problems resulting from two of the tasks required of such operators. The first task is the tracking of one of several simultaneously presented messages (selective listening, focused attention). The second is the tracking of several simultaneously presented messages or signals (divided attention).

3.1.1.1. Focused Attention

Selective listening or focused attention tasks require the operator (the listener) to focus on one of two or more simultaneously presented messages while disregarding the other(s). In such situations, the listener must separate the components of the wanted message from those of the unwanted background (other messages). The "cocktail party effect" (Cherry, 1966) is such an example. The "cocktail party effect" refers to the ability of party guests to attend to one conversation although many may be occurring, even if that conversation is more distant or less loud than others.

The effectiveness of selection has been studied in two ways: (1) by comparing the detection, recognition, and/or comprehension of auditory inputs presented alone with those presented under simultaneous listening conditions; and (2) assessing the effectiveness of ignoring/rejecting an auditory input. Further discussion describes research on the comprehension of messages and the effects of ignoring messages. A review of other research may be found in Hawkins and Presson (1986).

Research has identified a number of factors (cues) which influence selective listening performance: spatial location of the signal, pitch of the signal, semantic content of the signal/message, and intensity of the signal.

Spatial location (localization or lateralization) of a sound is determined in part by interaural time (phase) and interaural intensity (differences in time of arrival and intensity of the sound at the two ears). The importance of these cues in the performance of selective listening has been demonstrated by Licklider (1948). Licklider developed an improvement to a voice communications headphone set used by pilots. By altering incoming verbal messages so that the voice signal was out of phase at the two ears, yet leaving the external masking noise (noisy environment) in the same phase at the two ears, pilots were better able to avoid the masking effect of the external noise. This reduction in masking associated with the separation of the apparent source locations of the signal and noise is called the masking level difference.

Spatial separation of message sources (free-field) or of the auditory images of messages (headphones) has been used to reduce both the masking and confusability associated with presenting simultaneous messages. Spieth, Curtis, and Webster (1954) investigated the effects of speaker separation under free-field listening conditions. Subjects were presented with two simultaneous messages each spoken by a different voice of the same sex. Messages contained a code name, the name of the channel calling, the number of the caller, and a question about the visual display in front of the subject. The subject's task was to report the channel and talker calling, and answer the question posed. In some conditions, the subject had the option to switch the message to a nearer speaker or to the headphones. The degree of speaker separation (0 degrees, 10 to 20 degrees, or 90 to 180 degrees) was maintained in these conditions. In other conditions, visual cues (indicator lights specifying the relevant channel) were used. Differential frequency filtering of the messages, alone or in combination with the visual cues, comprised a final set of conditions. Results of the experiment showed that performance (correct channel identifications and correct answers) improves with speaker separation except under conditions where the task is already quite easy (i.e., visual cues and filtering are present).

The experiment by Spieth, Curtis, and Webster (1954) also provides an example of how pitch can be used as a cue in selective listening. All message pairs used in this experiment contained both a relevant and a distracting message. High pass (all frequencies above a given level can be heard) and low pass (all frequencies below a given level can be heard) filtering of the messages was used to create seven dual-message listening conditions. Results indicated that filtering significantly enhanced performance especially in conditions where no other cues were present (i.e., spatial separation). This implies that until the point at which filtering begins to impair the intelligibility of a relevant message, procedures which enhance the distance between the frequency bands of relevant and distracting signals will aid in selective listening.

Semantic content has also been studied as a cue in selective listening (Broadbent and Gregory, 1964; Miller and Selfridge, 1950; Treisman, 1964). Results of this research suggest that the semantic structure of auditory messages is as useful a cue for selective listening as is spatial separation. Selection performance may, therefore, be improved by providing semantic differences between relevant and irrelevant information.

Another factor which influences selective listening performance is the intensity of the auditory signal(s). The effectiveness of signal intensity (of both relevant and distracting messages) as a cue has been studied by Egan, Carterette, and Thwing (1954). In their experiment, subjects were presented with two simultaneous messages, either monaurally or dichotically. Each message began with a unique call sign. For example:

> LANGLEY BASE...next Tuesday we must vote...
> MITCHELL FIELD...the fur of cats goes by many names...

The subject's task was to reproduce all messages that followed the target call signs specified prior to the experiment. The results showed that as the relative intensity of the attended message was increased, selectivity improved. This improvement accrued more rapidly in conditions where the messages were presented dichotically than in those where they were presented monaurally.

3.1.1.2. Divided Attention

The listener's task in situations requiring divided attention is quite different from that of selective attention. Rather than attending to only one of several messages, the listener must now attend to two or more of those messages, responding to each as needed. A majority of the research on divided attention has attempted to determine the conditions under which attention can be successfully split between simultaneous inputs. This research suggests that in situations where the listener is monitoring two channels, yet is listening for a single target, no divided attention costs will occur. However, when listening for multiple, independent targets, task performance will depend on (1) the listener's perception of the events presented through channels other than the channel through which the target is presented, (2) the amount of practice the listener has had on the task, and (3) the modality in which the stimulus inputs are presented (auditory or auditory plus another modality).

The listener's reaction to events in the off-channel will be one of four types: (1) a hit (the off-channel carries a target that is correctly identified), (2) a miss (the off-channel carries a target which is not identified), (3) a false alarm (a nontarget in the off-channel is identified as a target), and (4) a correct rejection (a nontarget is correctly identified). In general, the listener's performance at identifying an input through a given channel is best when the response to the off-channel event is a correct rejection, intermediate when it is a miss, and poorest when it is a hit or false alarm.

To date, little research has investigated the effects of practice on divided attention. Ostry, Moray, and Marks (1976) found practice did improve performance on divided attention tasks. Performance in this study did not, however, increase to the level of focused attention performance.

The general conclusion of the research on divided attention and resource modalities suggests that strong divided attention effects occur with heteromodal stimulus presentation just as they do with homomodal presentation. Which type of presentation is superior remains unclear. Moray (1988) ties together the concepts of competing resources and extended practice. Although evidence suggests that resources are separate in the brain (i.e., visual and auditory processing do not generally compete for the same neural mechanisms), they are not completely independent. Some amount of interference (either in the form of delay or inaccuracy) will occur when attention is shared between sense modalities. With extended practice some aspects of processing can be automatized. This is especially true if the stimuli in question consistently requires the same response (i.e., consistent mapping). Automatization of these processes generally leads to improved performance.

3.1.2. Auditory Memory

Auditory or echoic memory is the memory system which stores the physical (acoustic) properties of auditory inputs. At the presentation of an auditory stimulus, this system stores a representation of that stimulus in a code very similar to the original input. The information is retained until it can be selectively processed and recorded into short-term memory.

A variety of factors affect the retention of information in auditory memory. A number of researchers have shown that the temporal interval between a memory item and an interfering stimulus will affect information retention. Hawkins and Presson (1977) and Massaro (1970) have found that as the delay between the presentation of a test stimulus (an auditory tone) and a masking stimulus increases, degradations (i.e., reduced recall accuracy of the test stimulus) in echoic memory retention decrease to near zero.

Laterality of the interfering stimulus item (i.e., masking tone) also appears to be a factor affecting the retention of information in echoic memory. The auditory system is extremely sensitive to differences in the timing and intensity of stimuli presented to both ears. Laterally separating the test and masking stimuli has been found to reduce performance degradations that are caused by the interference stimuli (Hawkins and Presson, 1977; Massaro, 1970).

3.2. MODELS OF INFORMATION PROCESSING

A large number of models of human information processing have been postulated. This section describes the major models of information processing upon which human performance and workload assessment techniques have been based.

Broadbent (1958) described a limited capacity filter model of human information processing. In Broadbent's model, the human may simultaneously receive input directly from any or all of the senses (parallel processing). This information is transmitted directly from the senses to some short-term storage area where a "selective filter" determines which information is to be processed through the limited capacity channel (central processing). It is at this point that parallel processing stops. The processing channel can now handle only one source of input at a time. Broadbent proposed this theory based upon his studies of dichotic listening. Noticing that subjects' comprehension of verbal messages decreased when subjects heard two different messages simultaneously (one in each ear), Broadbent proposed that subjects have the capacity to listen to only one voice at a time.

This trade-off between attending to one voice or another may also be explained as a simultaneous sharing of attention rather than a switching back and forth (as described by Broadbent). Treisman (1964) described an attenuation theory of processing to explain Broadbent's results as well as her own. This theory suggests that subjects who were instructed to attend primarily to one of two voices would be able, at the same time, to allocate a small portion of their attention to the other voice. With this theory, Treisman was able to account for a subjects' apparent sensitivity to certain kinds of information presented to the nonattended ear (i.e., their own name).

Norman (1976) went further to suggest that the selection of information occurs not by selectively blocking or filtering sensory information, but by selectively processing information already evoked or activated in memory by incoming sensory information.

These theories and others (Cherry, 1953; Moray, 1959) came to be known as "bottleneck" models of human information processing. They have in common that they seek to determine at what stage of processing a parallel system (capable of processing separate channels concurrently) narrows to a serial system that can handle only one channel at a time. Bottleneck theories can be divided into two general classes: early-selection theories (Broadbent, 1958; Treisman, 1969) that consider the bottleneck to occur at perception, and late-selection theories (Deutsch and Deutsch, 1963; Norman, 1968) that consider the bottleneck to occur at the point at which decisions to initiate responses (i.e., storage of information in long term memory, rehearsal of information) are made.

3.2.1. Capacity Theories

Capacity theories of human processing came about as a direct result of human factors research on mental workload. Capacity theories conceptualize the human as possessing a "pool" of processing facilities or "resources." The concept of resources will be described in greater detail later in this section. The first capacity theory was presented by Knowles (1963). Knowles theory, having direct application to operator task performance, proposes that as a task (primary task) demands more of an operator's resources (i.e., becomes more difficult), fewer of those resources are available for successful concurrent performance of a second task (secondary task). Performance of this second task, then, is expected to deteriorate.

The major distinction between bottleneck theories and capacity theories is that, rather than the structures of the human processing system being dedicated only to one task at a time, capacity may be allocated in varying amounts among separate activities (i.e., numerous tasks). Many other researchers have made contributions to the further development of capacity theory (Moray, 1967; Moray, Johannsen, Pew, Rasmussen, Sanders, and Wickens, 1979). This research allowed the resource metaphor to develop from a concept into a quantitative theory with testable predictions and important implications for the measurement of human behavior.

Perhaps the most important recent contribution to capacity theory has been made by Wickens (1984b). Whereas other capacity theories have assumed a single pool of undifferentiated resources available to all stages of processing, Wickens proposed the existence of multiple resources. According to this view, the human information processing system contains a number of commodities which may be assigned "resource-like" properties (i.e., sharing, allocation). The major components of multiple-resource theory will now be discussed.

3.2.2. <u>Resources</u>

The concept of resources may be loosely defined as processing facilities existing in some finite amount (Navon and Gopher, 1979). Other researchers have referred to this concept as effort, capacity, and attention (Kahneman, 1973; Moray, 1967; Shiffrin, 1976). Multiple-resource theory assumes that these resources reside in separate "reservoirs" or "pools" (Figure 4). This is contrary to single-resource theory which assumes one undifferentiated, shareable pool of resources.

Wickens (1980, 1984b) has developed a framework for determining the functional composition of these attentional resource reservoirs based upon the results of a large number of dual-task studies. This framework defines an operator's resources as a three-dimensional metric consisting of stages of processing (perceptual/central processing versus response), codes of perceptual and central processing (verbal versus spatial), modalities of input (visual versus auditory), and modalities of response (manual versus vocal) (Figure 5).

Due to the independent, nonoverlapping characteristics of Wickens' concept of resources, a number of implications follow: (1) tasks demanding completely nonoverlapping resources will always be perfectly time-shared, and (2) if two tasks utilize partially separate resources, the degree of interference (time-sharing efficiency) will be unaffected by the distance between the nonoverlapping components of these resources. Wickens (1984b) provides data supporting these implications.

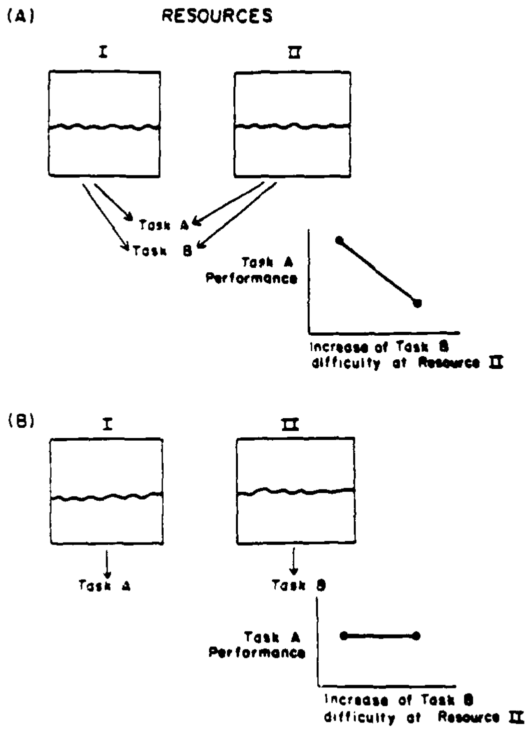

Figure 4. Difficulty Performance Trade-Offs [(A) Task A and B Share Resources I and II, (B) Tasks A and B Demand Exclusively Resources I and II] (Wickens, 1984b)

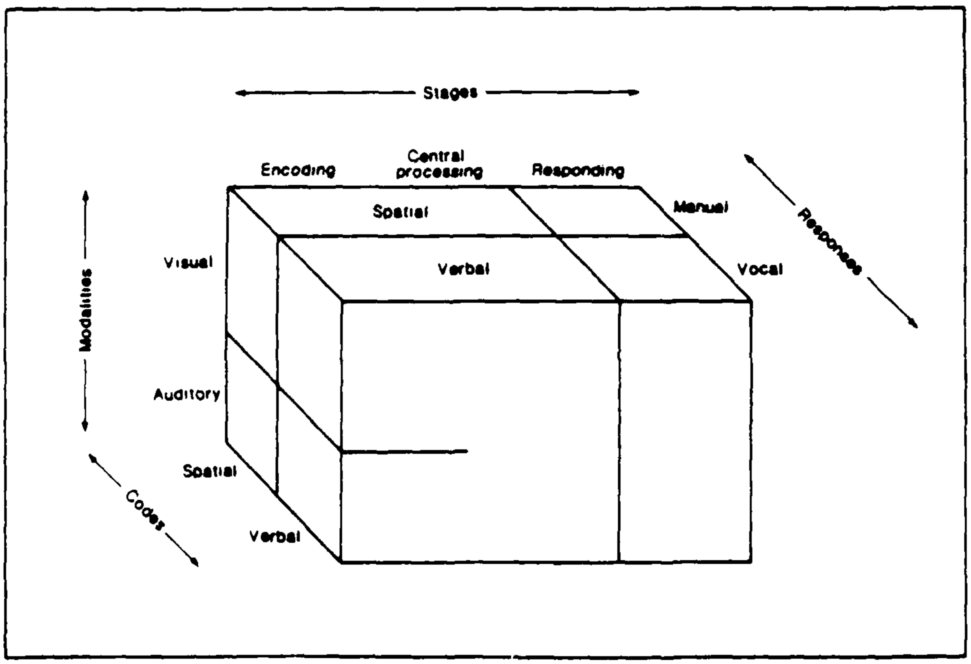

Figure 5. Resource Modalities (Wickens, 1984b)

The usefulness of the multiple resource concept to the applied science community is that it allows the researcher or designer to predict what combinations of task components have the potential to cause poor operator performance. This information may then be used to reevaluate and redesign equipment, tasks, or strategies which will result in optimal operator performance.

Section 4
OPERATOR PERFORMANCE MEASURES

The human performance literature contains a large number of theories and techniques for measurement of human performance in a variety of situations. Because the nature of AAMRL/BBA's current research on the effectiveness of voice communication in interactive environments involves such a strong cognitive component, SRL has chosen to concentrate a large portion of its literature survey efforts in the area of mental workload. Use of the mental workload literature, as opposed to some of the other areas of the human performance literature, has a number of advantages. First, mental workload research is founded on the basic principles of psychology and physiology. Both theoretical and applied research in this area have utilized basic principles in human information processing, cognition, learning and memory, arousal, motivation, etc. Second, the majority of research on mental workload has been conducted for eventual application in operational environments, specifically the flight environment. Finally, the mental workload literature contains well-defined guidelines for application of its theory and its tasks to specific situations (both experimental and operational). These guidelines are based upon well-documented, empirical evidence.

The following sections contain a review of a large amount of the mental workload literature, as well as some of the more basic literature on human performance. This review should not, however, be viewed specifically with respect to the assessment of workload. Rather, it should be viewed as a useful framework with which to view more basic research on human performance.

4.1. THE CONCEPT OF WORKLOAD

Mental workload has been defined in the literature in a broad variety of ways. Cooper and Harper (1969), for example, define workload as "the integrated physical and mental effort required to perform a specified task." Tennstedt (as cited in Roscoe, 1978) defines workload as "as summation of such processes as perception, evaluation, decision making, and actions taken to accommodate those needs generated by influences originating within or without the system." Many other authors have defined workload in many other ways. However, the majority of current definitions (Hart, 1982; O'Donnell and Eggemeier, 1983; Roscoe, 1978) define workload as being comprised of the following contributing factors: task demands, operator variables, and operator response (outcome). Task demands include such factors as difficulty, time constraints, time pressure, and criticality., Operator variables include effort, motivation, skill, experience, stress, personality, and fatigue. Finally, response considerations include mode of response (i.e., manual, verbal, simple, complex), feedback, and success. Figure 6 illustrates these factors and the ways in which they may combine to produce workload. Figure 7 illustrates a conceptual framework through which to

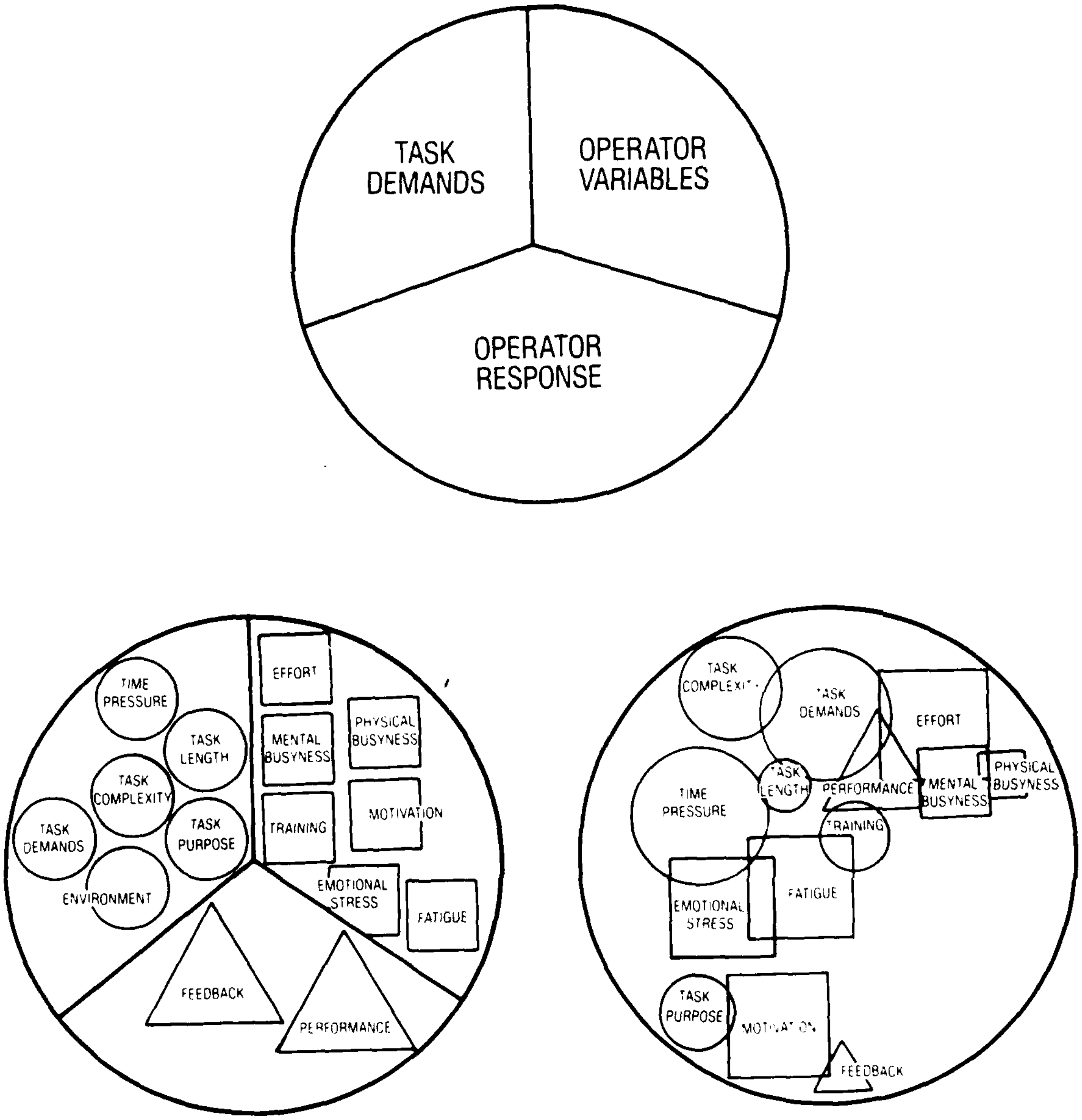

Figure 6. The Factors Comprising Workload

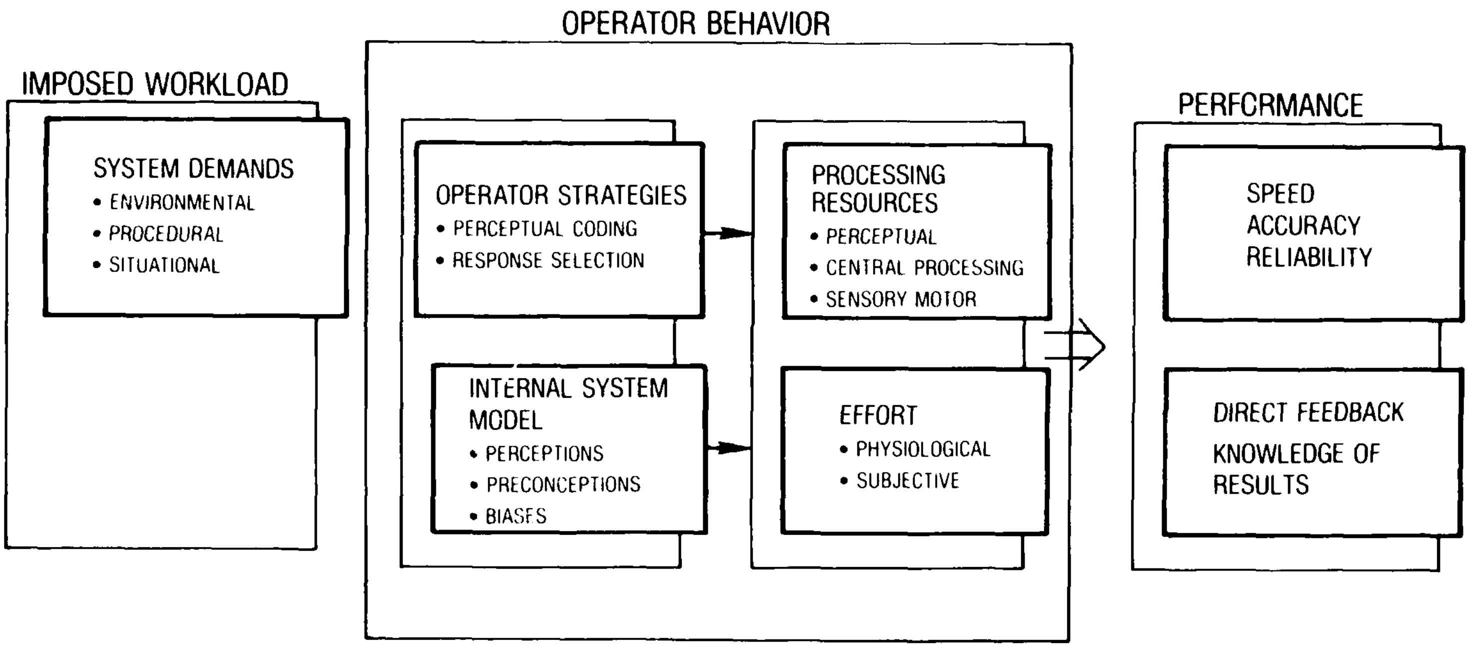

Figure 7. A Conceptual Framework

view the various workload factors and the ways in which they interact to affect performance. Thorough considerations of the concept of workload can be found in O'Donnell and Eggemeier (1983), Moray (1979), Roscoe (1978), and Wickens (1984a). For purposes of this report, however, workload will be defined as a multifaceted concept formed by the interactions of the demands of the task(s), operator effort, and performance outcome.

4.2. PERFORMANCE ASSESSMENT METHODS

4.2.1. Metric Selection Criteria

A number of criteria to assist in the selection of workload and other operator performance measures have been discussed in the literature. Because the majority of operator performance measures carry with them a number of intrinsic constraints and methodological requirements, thereby limiting their applicability, consideration of these selection criteria is important. This section will review the criteria to be used when selecting operator performance techniques for particular applications. The criteria to be discussed include sensitivity, diagnosticity, intrusion, implementation requirements, and operator acceptance.

- Sensitivity: Sensitivity is the term used in the literature to describe the ability of a measurement technique to detect changes in operator load that are caused by performance of a task or group of tasks (Chiles, 1982; Wickens, 1984a). Workload techniques, in particular, have been found to differ in sensitivity (Bell, 1978; Hicks and Wierwille, 1979; Wickens and Yeh, 1983; and Wierwille and Casali, 1983a). It is important to match the sensitivity of a technique with the requirements of an application. In some situations, a relatively insensitive technique may be sufficient, for example, if it is required only to identify areas of extreme workload in a system or procedure. Other applications may require finer discriminations of load (for example, when determining crew compositions). Choosing the sensitivity of a measurement technique is determined by the objective or goal to be satisfied by the use of that technique. If the objective is to determine whether a task or system already contains levels of loading which could lead to degraded operator performance, primary task measurement techniques will be adequate. If the goal is to determine whether or not the potential for overload and degraded performance exists, more sensitive techniques (physiological, secondary task, and subjective) should be used.

- Diagnosticity: The criteria of diagnosticity is based upon the multiple resources theory of capacity limitations of the human information processing system (see Section 3). Diagnosticity describes the ability of a measurement technique to discriminate the amount of task load imposed upon the different cognitive resources of the operator (e.g., perceptual versus central processing). Techniques have been found to differ in their degree of diagnosticity (Reid, Shingledecker, and Eggemeier, 1981; Wickens and Derrick, 1981; Wierwille and Casali, 1983b).

The diagnosticity of specific measurement techniques will be discussed further in later sections; however, subjective and primary task measures have generally been shown to exhibit low diagnosticity, while secondary and physiological measures are considered highly diagnostic. As with sensitivity, the choice between using a diagnostic versus a global measurement technique should be determined by the objective to be met. If the goal of the research effort is simply to determine if a loading problem exists somewhere in the system, techniques associated with low diagnosticity (i.e., subjective, primary task) will be adequate. If specific information concerning the locus of a previously identified problem (i.e., to suggest design modifications) is desired, more diagnostic techniques (secondary task, physiological) should be chosen.

- Intrusion: Intrusion refers to the degree to which a measurement technique degrades ongoing primary task performance. Certain degrees of primary task intrusion may be acceptable in some situations. In laboratory or simulation applications, intrusion may not be a great consideration. However, in many operational applications, due to safety considerations, the use of techniques which might cause degradations in primary task performance is precluded. Intrusion can also cause problems in data interpretation. Measurement techniques which cause significant degradations in primary task performance cannot be used to accurately predict the amount of load required for unimpaired performance on the primary task. The degree of intrusion associated with the various operator performance tasks again appears to differ. Although the database addressing this issue is small (see Casali and Wierwille, 1982; Ogden, Levine, and Eisner, 1979, Rolfe, 1971; Wierwille and Connor, 1983), it appears that subjective and physiological techniques are associated less with problems of intrusion than are secondary task techniques.

4.2.2. Major Classes of Assessment Techniques

Currently, there exist three major classes of human performance and workload measurement techniques: physiological techniques, subjective techniques, and performance-based techniques.

4.2.2.1. Physiological Techniques

The rationale for using physiological measures to study aspects of human performance such as workload is based upon the concept of "activation" or "arousal" (Roscoe, 1978). Arousal can be defined as "a state of preparedness of the body associated with increased activity in the nervous system" (Roscoe, 1978). It is assumed that arousal and performance are directly related, so that varying levels of physiological activity should provide realistic estimates of differing levels of workload or performance. Implicit in this assumption is the need to monitor not only physiological activity, but performance as well.

The overall usefulness of physiological techniques as measures of operator performance is unclear. O'Donnell and Eggemeier (1986) argue that, although the concept of measuring workload through physiological processes would seem simple, a majority of such efforts have failed to find consistent patterns of physiological change to correspond with known changes in workload. Hassett (1978) has suggested that, rather than viewing physiological measures as global indices of effort, arousal, or activation, they should be viewed instead as potential indices of specific psychological processes.

The following subsections discuss some of the more commonly used physiological measures of mental workload and human performance. These measures will be discussed only briefly; further detail can be found in O'Donnell (1979).

4.2.2.1.1. Measures of Brain Function

The electroencephalogram (EEG) records the brain's activity via surface electrodes placed directly on the scalp. Such measures have been taken during and after the performance of a task in hopes that the overall activation level in the brain would change directly as a function of imposed task load. Such techniques have not, however, yielded consistent or interpretable results (Lawrence, 1979; O'Donnell and Eggemeier, 1986; O'Donnell and Wilson, 1987; Roscoe, 1978). Other measures of brain function, such as various measures of evoked cortical response have shown more impressive results: signal analysis techniques (Callaway, Tueting, and Koslow, 1978); transient cortical evoked response (Lawrence, 1979; O'Donnell, 1979; Squires, Wickens, Squires, and Donchin, 1976); transient response to primary task (Gomer, Spicuzza, and O'Donnell, 1976); steady state evoked response (Reagan, 1977); and multiple site recording (Doyle, Ornstein, and Galin, 1974; Gevins, 1983). These measures appear to be useful for assessing the performance effects of task load.

4.2.2.1.2. Measures of Eye Function

Measures of eye function are valuable methods for task performance assessment because of their low intrusiveness, high operator acceptance, and ease of implementation. The most frequently used measures include pupillary response, eye fixation, scanning patterns, eye blinks, and movement speed. These measures have generally yielded consistent and sensitive results; however, they are relatively undiagnostic, providing only very global indications of task load.

4.2.2.1.3. Measures of Cardiac Function

The electrocardiogram (ECG), blood pressure, blood volume, and oxygen concentration have all been used as cardiac indicators of overall workload and specific task load (O'Donnell and Eggemeier, 1986). Although measures of cardiac function have been somewhat successful as predictors of workload in several studies (Casali and Wierwille, 1983; Hicks and Wierwille, 1979; Wierwille and Connor, 1983), it is unclear exactly how cardiac function changes with different types and amounts of task load (O'Donnell and Eggemeier, 1986). Until more data are established, these measures must be considered potentially useful but unvalidated measures of task load.

4.2.2.1.4. Measures of Muscle Function

Myoelectric signals generated by muscle contractions have also been used to measure mental and physical workload using an electromyograph (EMG). These signals are measured either with surface electrodes placed directly over the muscle, or needle electrodes placed directly into the muscle. Physical work is indicated by the actual muscle activity at the specified muscle, while mental work is indicated by the static tension level of a muscle not directly involved in the performance of the task (O'Donnell and Eggemeier, 1986). Current measures of muscle function, due to the necessities of their measurement techniques, are not recommended, as they are not simple, sensitive, or diagnostic, and have obvious intrusion and safety limitations.

4.2.2.2. Subjective Techniques

Subjective measures of operator effort and task load require the operator to report the amount of "load" experienced in the performance of a particular task or set of tasks. The majority of such techniques described in the literature are designed specifically to assess "workload," rather than simple "task load." However, as the two concepts are similar (task load may be viewed as one component of workload), subjective workload assessment techniques will be described and discussed with the assumption that they are also applicable to task loading situations.

Subjective measures have been used extensively to assess operator workload due to their practical advantages (ease of implementation, nonintrusiveness, high operator acceptance), and their capability of discriminating among different levels of load (sensitivity) (Moray, 1982; O'Donnell and

Eggemeier, 1986; Williges and Wierwille, 1979). Subjective measures are not, however, considered diagnostic (O'Donnell and Eggemeier, 1986). Available evidence (O'Donnell and Eggemeier, 1986) suggests current measures represent a global measure of load and, therefore, should be used as general screening devices to determine if overload exists anywhere within task performance. The most commonly used rating scales and psychometric subjective workload assessment techniques will now be described.

4.2.2.2.1. Rating Scales

The Cooper-Harper Aircraft Handling Characteristics Scale (Cooper and Harper, 1969), designed for use by test pilots to assess the ease of control of various aircraft, has been used extensively as an index of mental workload. This ten-point rating scale requires the pilot to judge the adequacy of an aircraft for some specified task or operation. The assumption that handling qualities and operator workload are directly related has been supported by a number of research efforts (Hess, 1977; Moray, 1982; Williges and Wierwille, 1979).

Modified Cooper-Harper rating scales (North, Stackhouse, and Graffunder, 1979; Wierwille and Casali, 1983a; Wolfe 1978) have also been used to measure mental workload. These scales are quite similar to the original Cooper-Harper scale, with the exception that references to aircraft handling characteristics in the original scale were replaced by descriptors of pilot workload effort. As with the original Cooper-Harper scale, available data support its sensitivity to varying levels of load but again suggest its lack of diagnosticity (North et al., 1979; Wierwille and Casali, 1983b; Wolfe, 1978).

Two other rating scales have been used to measure factors associated with workload. These scales, generally known as University of Stockholm Scales, measure the perceived difficulty and perceived effort of the operator. These ratings scales have been used in conjunction with intelligence test items (reasoning, spatial ability, and verbal comprehension) (Bratfisch, 1972; Bratfisch, Borg, and Dornic, 1972; Hallsten and Borg, 1975), visual discrimination tasks, letter transformation tasks, digit transformation tasks, and visual auditory detection tasks (Bratfisch, Borg, and Dornic, 1972; Dornic, 1980).

Overall, rating scales used as subjective measures of mental load have proven to be sensitive indicators of operator effort and expenditure. They are nonintrusive, are easily implemented, appear to have high operator acceptance, and are generally not time-consuming. Rating scale measurements are, however, relatively undiagnostic and should be interpreted as global indicators of operator mental load.

4.2.2.2.2. Interviews and Questionnaires

Interviews and questionnaires have also been used as techniques to gather subjective data on operator mental load. Williges and Wierwille (1979) describe the variety of these procedures which range from open-ended debriefing sessions to carefully designed questionnaires. As these techniques are less structured than rating scales, obtained data may be difficult to interpret. These techniques can, however, be valuable when used in conjunction with other measures, by providing information which might not otherwise be obtainable. Again, it is recommended that questionnaires and interviews not be relied upon as stand-alone techniques for assessing operator load.

4.2.2.2.3. Psychometric Techniques

Psychometric measures which have been used to assess operator load include magnitude estimation, paired comparison, equal-appearing intervals, and conjoint measurement. These methods can generate interval-scaled data which provide certain interpretation advantages in data analysis over many of the other subjective assessment techniques. Detailed descriptions of these techniques are provided in O'Donnell and Eggemeier (1986).

4.2.2.3. Performance Based Measures

Performance based measures derive an index of operator loading from some aspect of operator behavior or activity (i.e., task performance). These measures are also referred to as behavioral measures (Shingledecker, 1983; Williges and Wierwille, 1979).

4.2.2.3.1. Primary Task Measures

Primary task methods measure the operator's performance on some task or design option of interest. It is assumed that, as the load on the operator increases, performance of the task will change, usually resulting in some amount of degradation. Measurement of that degradation is used to provide an index of the load associated with the task. The workload literature describes two types of primary task measures: single and multiple primary task measures.

- Single Primary Task Measures: Single primary task measures use a single aspect of primary task performance (number of errors, speed of performance, etc.) as an indication of operator load. In this paradigm, the primary task measure should be chosen to reflect an aspect of performance that is expected to be influenced by the manipulation of the load. This is often difficult. However, it is a critical consideration as the success of the evaluation is dependent on a single parameter of performance.

Many successful applications of this paradigm are described in the literature (Hicks and Wierwille, 1979; Isreal, Chesney, Wickens, and Donchin, 1980; Kraus and Roscoe, 1972; Wierwille and Connor, 1983; Williges and Wierwille, 1979). Single primary task measures have successfully distinguished variations in load, especially across moderate levels of load, as well as discriminating overload from nonoverload situations.

Instances in which appropriate single primary task measures have failed to reflect manipulations of task load have also been reported (Bell, 1978; Burke, Gilson, and Jagacinski, 1980; Schultz, Newell, and Whitbeck, 1970).

- Multiple Primary Task Measures: Data can also be collected on multiple aspects of a primary task. This paradigm is generally used in simulated or real-world environments or in complex laboratory task situations. Generally, both error and latency data are gathered for several dependent variables (DV). The assumed advantage in using multiple primary task measures is that they will provide greater sensitivity to changes in operator load by: (1) decreasing measurement error via combined data analysis of multiple DVs, and (2) allowing for the assessment of more than one resource or skill, thereby increasing the precision and utility of the measure. Although the selection of task parameters for this methodology is not as critical as for the single task methodology, parameters to be measured should again be chosen based upon their potential to be influenced by different load manipulations. O'Donnell and Eggemeier (1986) caution that this is an important consideration, as data collected simply because of availability may not be meaningful.

Multiple primary task measures, as with single task measures, have produced mixed results as to their capability to distinguish among different levels of load. A number of experiments (Dorfman and Goldstein, 1975; Goldstein and Dorfman, 1978; Hicks and Wierwille, 1979) have found multiple primary task measures sensitive to variations in load. Others (Brecht, 1977; Finkelman, Zeitlin, Filippi, and Friend, 1977) have found that some measures fail to discriminate variations in load that were detected by other assessment techniques. Again, although multiple primary task measures may, in certain instances, discriminate overload from nonoverload situations, they generally do not provide clear diagnostic statements as to the specific resources being overloaded. Therefore, multiple primary task measures, like single task primary measures, should be regarded as global measures of operator load.

The principal utility of primary task performance measures is in determining whether the load associated with a system (task, equipment, environment) will degrade operator performance. In such applications, where diagnostic capability is not required, either single or multiple summary task measures will provide adequate information.

4.2.2.3.2. Secondary Task Measures

Secondary task measures of operator load require the concurrent performance of two tasks by the operator. The task of central interest is generally termed the "primary task," while the additional task is termed the "secondary task." The estimate of operator load will be obtained from the operator's performance on the secondary task. Secondary task methodology may be used in a variety of applications, including the measurement of operator effort, attentional demand, the effect of stressors, and the adequacy of displays. It is most often used as a measure of the spare or residual capacity of the operator as he performs the primary task (see Section 3). Secondary task methodologies have proven to be both sensitive to differences in capacity expenditure and diagnostic of primary task demand (they are capable of discriminating some differences in resource expenditure, i.e., central processing versus motor output). Some intrusion (the degree to which the secondary task degrades primary task performance) problems have been reported (Ogden et al., 1979; Williges and Wierwille, 1979). In attempts to alleviate the intrusion problem, several techniques (e.g., embedded secondary task, adaptive procedures) have been designed. These techniques are reviewed in O'Donnell and Eggemeier (1986).

In most applications of secondary task measures, the operator is instructed to maintain error-free performance on one task at the expense of the other. Depending upon the goal of the experimenter, one of two different methodologies may be used: the loading task paradigm or the subsidiary task paradigm. The loading task paradigm instructs the operator to maintain a certain level of performance on the secondary task, even if decrements in primary task performance result. The assumption of this paradigm is that the additional load imposed on the operator by the secondary task will shift total operator load from Region A to Region B (Figure 8), inducing degradations in performance of the primary task. When levels of secondary task load are equal, performance degradation will be greater for difficult primary tasks than for easy primary tasks. Degradations in primary task performance that occur at specific levels of secondary task load are used as an index of the resulting cognitive load (workload) associated with the primary task. Secondary task performance is measured directly to ensure that the specified criteria levels are maintained and that the load imposed by the task is equated across the various experimental conditions.

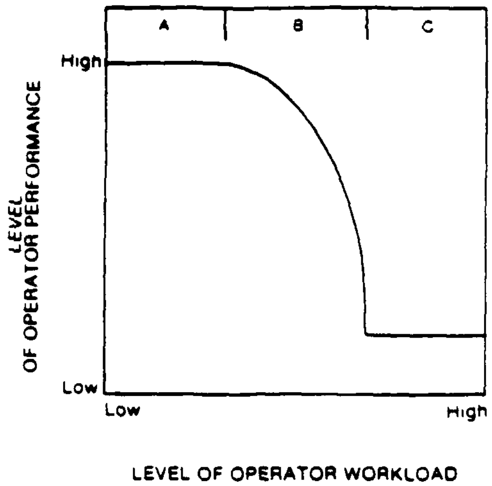

Figure 8. Hypothetical Relationship Between Workload and Operator Performance (O'Donnell and Eggemeier, 1986)

The loading task paradigm has been used primarily to simulate the effects of information processing demands that are absent from the laboratory, but are expected to occur in the operational environment. Dougherty, Emery, and Curtin (1964), for example, used a loading task paradigm to evaluate two cockpit display options (conventional versus pictoral). Primary task measures had previously indicated no differences in the cognitive load imposed upon pilots by the two displays. Addition of a secondary digit naming task, however, caused decrements in flight performance (the primary task) under the traditional display condition. Since equivalent levels of secondary task load did not lead to performance decrements in the pictorial display condition, it can be concluded that the pictorial display imposed less load on the pilot than did the conventional display. This example illustrates the ability of secondary loading tasks to increase the operator's load in the laboratory, making it more representative of the operational environment and increasing the sensitivity of primary task measures. Other applications of the loading task paradigm have included the evaluation of methods of task performance, the evaluation of display configurations, and the effects of stressors (noise, heat) on primary task performance (Ogden et al., 1979; Rolfe, 1971).

The second secondary task paradigm (i.e., the subsidiary or reserve capacity task paradigm) is more frequently used. In this paradigm, the subject is instructed to avoid degraded primary task performance at the expense of the secondary task. The secondary task, rather than being used to load the primary task (as in the loading task paradigm), is now used to determine how much additional work the operator may do while performing the primary task at some specified level (its single task baseline level). The assumption of the subsidiary task paradigm is that, as the second task is added, decrements in that task will result (again, as measured against its single task baseline level). These decrements will then serve as a measure of the reserve capacity of the operator when performing the primary task (Brown, 1964; Knowles, 1963). The subsidiary task paradigm has been used to measure reserve capacity for a variety of purposes, including evaluation of instruments and displays, operating conditions and procedures, and the effects of extended practice on performance (Ogden et al., 1979; Rolfe, 1971; Williges and Wierwille, 1979).

4.3. METHODOLOGICAL CONSIDERATIONS FOR SECONDARY TASK TECHNIQUES

When using secondary task techniques to measure operator load, there are a number of methodological guidelines to be considered. A number of general guidelines for the use of secondary task techniques have already been discussed in previous sections. These guidelines are shown again in Table 2. A thorough review of these guidelines including specific techniques to minimize primary task intrusion, techniques to ensure secondary task sensitivity, interpretation of single-to-dual task performance decrements, and the most frequently used types of secondary tasks can be found in O'Donnell and Eggemeier (1986).

TABLE 2. METHODOLOGICAL GUIDELINES FOR APPLICATIONS OF SECONDARY TASK METHODOLOGY (O'Donnell and Eggemeier, 1986)

1. In the loading task paradigm, subjects should be instructed to maintain secondary task performance at single-task baselines under concurrent task conditions.
2. In the subsidiary task paradigm, subjects should be instructed that primary task performance should be maintained at single-task baseline levels under concurrent task conditions.
3. In both paradigms, baseline measures of single-task performance on both the primary and secondary tasks should be taken. In the loading task paradigm, primary task baselines are required to assess differences in primary task performance that might occur under concurrent task conditions. Secondary task baselines are required to ensure that the secondary task is performed to the criterion set by the experimenter. In the subsidiary task paradigm, primary task baseline performance is required to evaluate any intrusion effects that might occur. Baseline secondary task measures are required to evaluate properly the degree of single to dual task decrements which might occur.
4. In both paradigms, employ several levels of secondary task difficulty. Higher levels of secondary task difficulty may distinguish differences in workload between design options or tasks that are not distinguished by lower levels of secondary task difficulty. The theoretical basis for such difficulty effects is that lower levels of secondary task difficulty may not be sufficient to shift total workload from Region A to B (Figure 8), whereas more difficult levels may do so.
5. In the subsidiary task paradigm, consider the use of various techniques that have been proposed to reduce or eleminate primary task intrusion. Two major classes of these techniques include adaptive secondary methodology and embedded secondary tasks.
6. In both paradigms, attempt to ensure maximum secondary task sensitivity through choice of an appropriate task and through use of sufficient practice to achieve stable performance on the secondary task prior to its use.

Section 5

EXISTING U.S. AIR FORCE WORKLOAD BATTERIES

5.1. BACKGROUND

Workload batteries are collections of a number of experimental tasks which can be used to investigate a variety of research issues or questions concerning human performance and workload. Each individual battery task may be used alone or in conjunction with the other tasks in the battery. The utility of a workload battery is that it can be used to provide both global and diagnostic information. By using a specified task or set of tasks in the battery, the researcher may receive a global assessment of a particular situation or a general answer to a specific research question (i.e., "Is there a significant amount of workload associated with this system?"). The Unified Tri-Services Cognitive Performance Assessment Battery (UTC-PAB), for example, specifies a set of five tasks (each task representative of one of the major human information processing functions) to be used for initial global screening. Based upon the results of the initial screening, other tasks are specified by UTC-PAB for use in further, more diagnostic investigation.

In addition to the various types of information which can be obtained from workload batteries, these batteries have a number of methodological advantages. Existing batteries contain well documented experimenter instructions, subject instructions, and guidelines for use of the various tasks and task sets. Some existing batteries have been implemented in user-friendly microcomputer software. This software aids in both data collection and analysis. Finally, existing batteries contain documentation of the sensitivity and reliability of their various tasks.

Since at least 1980, AAMRL has been developing and collecting different tasks for compilation into test batteries (Eggemeier, 1981). Two major test batteries that have been developed are the Criterion Task Set (CTS) by Shingledecker (1984), and the UTC-PAB (Perez, Masline, Ramsey, and Urban, 1987). The CTS was designed to place selective demands on the basic mental resources and information processing functions of the subjects. The UTC-PAB was designed to evaluate cognitive performances of test subjects.

5.1.1. Criterion Task Set

The CTS was developed as a research tool for applied experimentation of human performance capabilities. The test battery is made up of nine standardized tasks. Eight of the nine tests can be presented with three different levels of difficulty. The single level test is a finger tapping test, designated as the Interval Production Task. All of the tasks are on the following list:

1. Probability Monitoring
2. Continuous Recall
3. Memory Search
4. Linguistic Processing
5. Mathematical Processing
6. Spatial Processing
7. Grammatical Reasoning
8. Unstable Tracking
9. Interval Production

All of these tests are implemented in user-friendly software on an inexpensive microcomputer system. A user's guide has been developed to provide information on: (1) system hardware, (2) system assembly, (3) data collection, and (4) data analysis (Acton and Crabtree, 1985).

5.1.2. Unified Tri-Services Cognitive Performance Assessment Battery

The UTC-PAB was developed to evaluate cognitive performance of subjects for a chemical defense biomedical drug screening program. The tests were selected by the Tri-Service Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD^3 MILPERF). A report by England, Reeves, Shingledecker, Thorne, Wilson, and Hegge (cited in Perez et al., 1987) details the history and selection criteria for the UTC-PAB. The test battery consists of 25 tests that evaluate six different cognitive processes. These cognitive areas and their associated tests are listed below:

1. PERCEPTUAL INPUT, DETECTION, AND IDENTIFICATION
   - Visual Scanning Task
   - Visual Probability Monitoring Task
   - Pattern Comparison (Simultaneous)
   - Four-Choice Serial Reaction Time

2. CENTRAL PROCESSING
   - Auditory Memory Search (Memory Search Tasks)
   - Continuous Recognition Task
   - Code Substitution Task
   - Visual Memory Search (Memory Search Tasks)
   - Item Order Test

3. INFORMATION INTEGRATION/MANIPULATION--LINGUISTIC/SYMBOLIC
   - Linguistic Processing Task
   - Two-Column Addition
   - Grammatical Reasoning (Symbolic)
   - Mathematical Processing Task
   - Grammatical Reasoning (Traditional)

4. INFORMATION INTEGRATION/MANIPULATION--SPATIAL MODE
   - Spatial Processing Task
   - Matching to Sample
   - Time Wall
   - Matrix Rotation Task (Spatial Processing Task)
   - Manikin Test
   - Pattern Comparison (Successive)

5. OUTPUT/RESPONSE EXECUTION
   - Interval Production Task
   - Unstable Tracking Task

6. SELECTIVE/DIVIDED ATTENTION
   - Dichotic Listening Task
   - Memory Search/Unstable Tracking Combination (Sternberg-Tracking Combination)
   - Stroop Test

A full description of the purpose, history, and utilizational instructions for each of these tests are reported by Perez et al. (1987). Like the CTS test battery, the UTC-PAB is also implemented on user-friendly software capable of running on an inexpensive microcomputer system.

Section 6
COMMUNICATION THEORY

Communication or information theory is a mathematical attempt to define the limitations of a specific communication system or process. The process of measuring communication can be divided into three subproblems (Weaver, 1949/1964); a "technical" problem, a "sematic" problem, and an "effectiveness" problem.

6.1. THE TECHNICAL PROBLEM

The technical problem of communication measurement is determining how accurately a set of symbols (i.e., written speech) or a signal (i.e., radio transmission of voice) is transferred from a sender to a receiver. Figure 9 represents the communication process at the technical level.

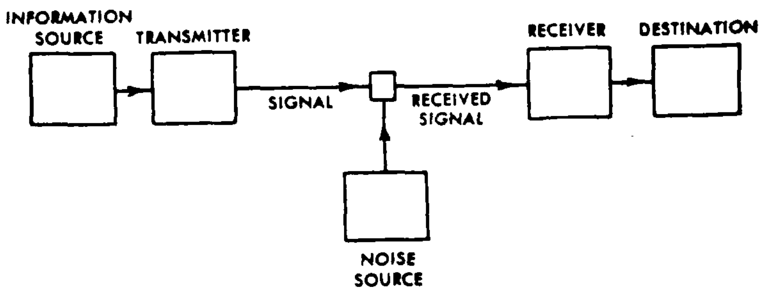

Figure 9. The Elements of a Communication System (Shannon and Weaver, 1949/1964)

This process may be described as follows. The information source (i.e., the sender) selects the desired message from a set of possible messages. This message is then converted by the transmitter into the signal. The signal is sent over the communication channel to the receiver. Finally, the receiver changes the transmitted signal back into a message and delivers it to the destination.

Shannon (1948/1964) developed a mathematical theory which describes communication at the technical level. Shannon's theory has been used to address a number of problems concerning communication systems, including: how to measure the amount of information within a system, how to measure the capacity of a communication channel, how to determine the characteristics of an efficient coding process, how noise affects the accuracy of receiving a message, how to minimize the undesirable effects of noise, and how continuous and discrete signals differently affect a communication system.

6.1.1. Information

In communication theory, the term "information" does not generally refer to its more ordinary definitions of "meaning," "knowledge," "news," etc. "Information" in this situation generally refers to the statistical rarity of a source of message symbols. Shannon (1948/1964) defined information as a "a measure of one's freedom of choice when one selects a message." In this sense, information describes not the content of individual messages themselves, but the amount of choice an individual has in selecting any particular message from the total message set. In other words, information is defined by the uncertainty of events--less certain events having a greater amount of information associated with them. Mathematically, the "self-information" of an event before or after transmission (given that the output is independent of the input) can be described as (Systems Research Laboratories [SRL], 1987):

$$(A = a_k) = I(a_k) = \log_2 [1/P\,(a_k)] \text{ bits}$$

where:

$A = (a_1, a_z, \ldots a_k)$, the set of inputs or the source alphabet/vocabulary.

$I(a_k)$ = the self-information of the event that $A = a_k$, or the information needed to make the occurrence of event a_k certain.

$P(a_k)$ = the probability that A was transmitted.

The selection of a message in a communication system can occur in a number of ways. In the simplest case (as described above), the information source is free to choose only between a few predetermined or "canned" messages. More commonly, however, the information source constructs each message individually, by making a sequence of choices from some set of symbols. An everyday example is choosing one word after another to form a sentence.

According to Weaver (1949/1964), the consideration of statistical probabilities becomes important for the measurement of communication because probabilities reflect the rules by which a message is formed. As each successive symbol from the vocabulary set is chosen, the probabilities of selecting the remaining symbols change. In other words, at any stage in the communication process, the probability of selecting any symbol is determined by the preceding choices. In English speech, for example, if the last selected symbol is "the," the probability that the next symbol is an article is very small, while the probability that it is either an adjective or a noun is very great.

Systems in which sequences of symbols are chosen according to probabilities are called "stochastic processes." Stochastic systems where probabilities depend directly upon the previous events are called "Markoff processes."

The ability of a communication channel to transmit information is described as its "capacity." Generally, capacity is defined as the amount of information transmitted per second, measured in bits per second.

6.1.2. Entropy

The entropy of a source alphabet or vocabulary is a measure of the randomness or uncertainty about that alphabet. Entropy can also be thought of as the average amount of information per source symbol. Mathematically, entropy is defined as (SRL, 1987):

$$H[I(a_k)] = \sum_{k=1}^{K} P(a_k) \log_2 [1/P(a_k)] \text{ bits}$$

where:

$H[I(a_k)]$ = the average amount of information per source symbol.

6.1.3. Mutual Information

"Mutual information" describes the uncertainty in some symbol or vocabulary item (i.e., a_k) that is resolved in the output of the system (i.e., b_j). The previously described measures of information have considered the output of a system independent of the input (i.e., information before or after transmission). However, in a real system, the output is dependent upon the input. The self-information of the event $A = a_k$, given that event $B = b_j$ has occurred, can be described as (SRL, 1987):

$$I(a_k/b_j) = \log_2 [1/P(a_k/b_j)] = -\log_2 [P(a_k/b_j)]$$

This describes the amount of information that must be supplied to an observer to specify that $A = a_k$ after an observer has received $B = b_j$ or, in other words, the amount of information that was lost during transmission. The difference between this quantity [i.e, $I(a_k/b_j)$] and the self-information of the event that $A = a_k$ [i.e., $I(a_k)$] is the mutual information. This is a measure of the gain in information due to the reception of a symbol (b_j). Mathematically, mutual information is defined as (SRL, 1987):

$$I(a_k; b_j) = I(a_k) - I(a_j/b_j) = \log_2 [P(a_k/b_j)/P(a_k)]$$

When the mutual information is averaged across the input alphabet or vocabulary, the "channel" or "average mutual information" (AMI) is obtained. This is a measure of the information gain of an entire system, not dependent on individual input and output symbols, but dependent on the symbol frequencies. The AMI is represented as (SRL, 1987):

$$I(A; B) = \sum_{k=1}^{K} \sum_{j=1}^{J} P(a_k, b_j) \log_2 [P(a_k, b_j)/P(a_k) P(b_j)]$$

6.1.4. Information Theory for Assessing Operator Performance

According to Wickens (1984a), a large amount of human performance theory is specifically concerned with the problem of transmitting information. In any situation where an operator is perceiving stimuli and somehow responding to that stimuli, the operator can be described as both encoding and transmitting information. As an example, Wickens (1984a) describes an aircraft pilot as someone who "must process a multitude of visual signals bearing on the status of the aircraft while listening to auditory messages from air traffic control concerning flight plans and the status of other aircraft." Information theory provides a metric that enables these information processes to be quantified and described in ways that allow the many tasks of the aircraft pilot (or other human operators) to be compared. When information theory is used in this way, it is assumed that information processing efficiency can be associated with the amount of information an operator can process per unit time (i.e., channel capacity), and that task difficulty can be associated with the rate of information presentation (Wickens, 1984a).

Information theory has been a great asset to researchers investigating both communication processes and operator performance. Wickens (1984a) states that information theory provides an essentially dimensionless unit of performance across a wide variety of different dependent variables. Fitts and Posner (1967) have also suggested that certain limits of the human information processing system remain relatively invariant when described in the terms of information theory. Despite these successes, however, the use of information theory in human performance research/ applications has received some criticism. Among these criticisms are limitations in the sensitivity of the information metric, and the inability of information measures to describe the factors influencing reaction time (RT). Wickens (1984a) offers a complete discussion of these criticisms.

Section 7
VOICE COMMUNICATION EFFECTIVENESS: PROPOSED TASKS

This section of this report describes the performance tasks selected for implementation in the PACRAT test facility. Section 7.1 describes the communications scenario that has been developed. Section 7.2 describes the secondary task selected to be performed concurrently with the communication scenario. Finally, Section 7.3 describes an alternate scenario configuration which could be further developed for use in the PACRAT facility.

7.1. COMMUNICATIONS SCENARIO

A communication scenario was developed by SRL for implementation in the PACRAT test facility. The scenario was constructed based upon the findings of the literature review and in accordance with a number of predetermined constraints (Figure 10). These constraints were determined after thorough consideration of both the characteristics of the PACRAT facility and AAMRL/BBA's research interests and requirements. The communication scenario is a 30-minute sequence of short, operationally realistic sentences which are verbally presented and require a series of manual responses by the subject (Figure 11). It models a two-way, interactive, time dependent communication situation. Each message is a separate, complete sentence, two to six words in length.

7.1.1. Selection of Vocabulary and Development of Message Sentences

The vocabulary used in the scenario was chosen from a combination of sets of confusable words previously developed by AAMRL/BBA and transcriptions of actual Air Force pilot communications. The confusable words were derived from 2,000 hours of Air Force in-flight communication in an attempt to develop a standardized word intelligibility test using flight jargon words. Ten lists of confusable words (Table 3) were selected for use. Each list consisted of 50 sets of four "confusable" words. The confusability of each word set had been previously determined by their acoustic and phonetic similarity, and by data collected from their experimental use. The sets resulted in a database (database A) of approximately 1900 words (some words appeared on more than one list or in more than one confusable set). This database was then combined with a database of approximately 2100 words (database B) selected from transcriptions of pilot communications during various flight situations. These transcriptions had been previously analyzed for entropy and mutual information (see Section 6). All words which were common to both databases (database A and database B) were combined to form a third database (database C). This database

- All Information Presented Over the Communication Channel
- Words Used in Messages Chosen from List of Confusable Words
- May Be Used Alone or with Another Task
- 30-Minute Duration
- Minimal Training Requirements
- Aircraft Oriented, but Understandable by Nonoperational Subjects
- Low Level Two-Way Interaction
- Easily Modifiable
- Structured Script
- Cost Associated with Message Repeats
- Cost Associated with Wrong Decision
- Forced to Make a Decision
- Time Constrained
- Get Through Message "Loop" Quickly

Figure 10. Scenario Constraints

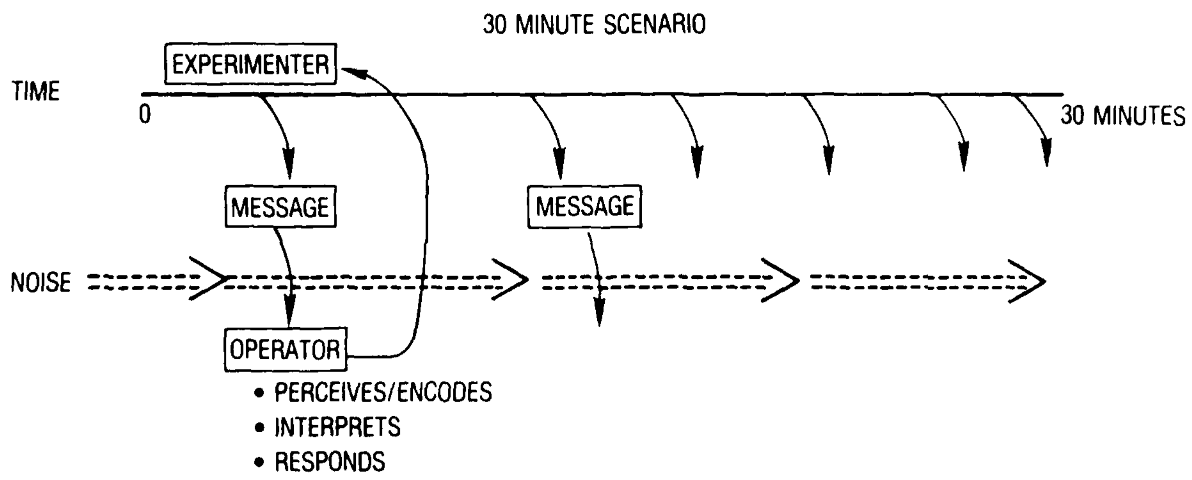

Figure 11. VCE Communication Scenario

TABLE 3. LIST OF CONFUSABLE WORDS

1.	Marked	Marsh	Marks	Mark
2.	Blast	Fast	Past	Last
3.	Cone	Code	Cove	Cold
4.	Reached	Reach	Reef	Reads
5.	Scan	Can	Span	Plan
6.	Seemed	Seals	Seems	Seized
7.	We	Free	Be	See
8.	Mapped	Match	Map	Matched
9.	Parts	Park	Parked	Part
10.	Real	She'll	Wheel	Feel
11.	Juts	Jump	Judge	Just
12.	Fire	Prior	Wire	Tire
13.	Great	Straight	Gate	State
14.	Thank	Bank	Rank	Yank
15.	Tight	Tied	Type	Timed
16.	Fake	Face	Failed	Phased
17.	Done	One	Gun	Ton
18.	Hack	Pack	Shack	Fac
19.	White	Right	Bright	Light
20.	Show	Though	So	Row
21.	Head	Held	Helps	Help
22.	Lose	Loose	Loop	Looped
23.	And	Add	Ask	Am
24.	She's	She'd	She	She'll
25.	Flap	Scrap	Cap	Slap
26.	Not	Dot	Shot	Hot
27.	With	Wing	Will	Width
28.	Did	Grid	Mid	Hid
29.	Six	Sixth	Sick	Sit
30.	Plate	Placed	Plane	Place
31.	Tripped	Trims	Trimmed	Trim
32.	Loud	Plowed	Cloud	How'd
33.	Than	Man	Can	Fan
34.	Or	For	Poor	Door
35.	Word	Heard	Bird	Third
36.	Old	Cold	Told	Hold
37.	West	Went	Wet	Well
38.	Be	Beam	Beached	Beach
39.	Notes	No	Note	Nose
40.	Wash	Washed	Watch	Watched
41.	Dust	Duck	Ducked	Ducks
42.	Heat	He'd	He's	Heats
43.	Notch	Knot	Knots	Notched
44.	Flown	Zone	Bone	Tone
45.	It	Its	Is	If
46.	Pick	Chick	Click	Quick
47.	Weak	Sneak	Seek	Peak
48.	Glide	Side	Slide	Guide
49.	Fifth	Fix	Fixed	Fits
50.	Tough	Buff	Rough	Stuff

consisted of approximately 1200 words. From this database, all words which had occurred as primary words (the first word in a set of two) in database B and had secondary words (the second word in a set of two) existing in database C were combined into a fourth and final database (data base D). This database of 496 confusable words was used as the vocabulary set for the scenario. Figure 12 illustrates the combination of the various databases into the final vocabulary.

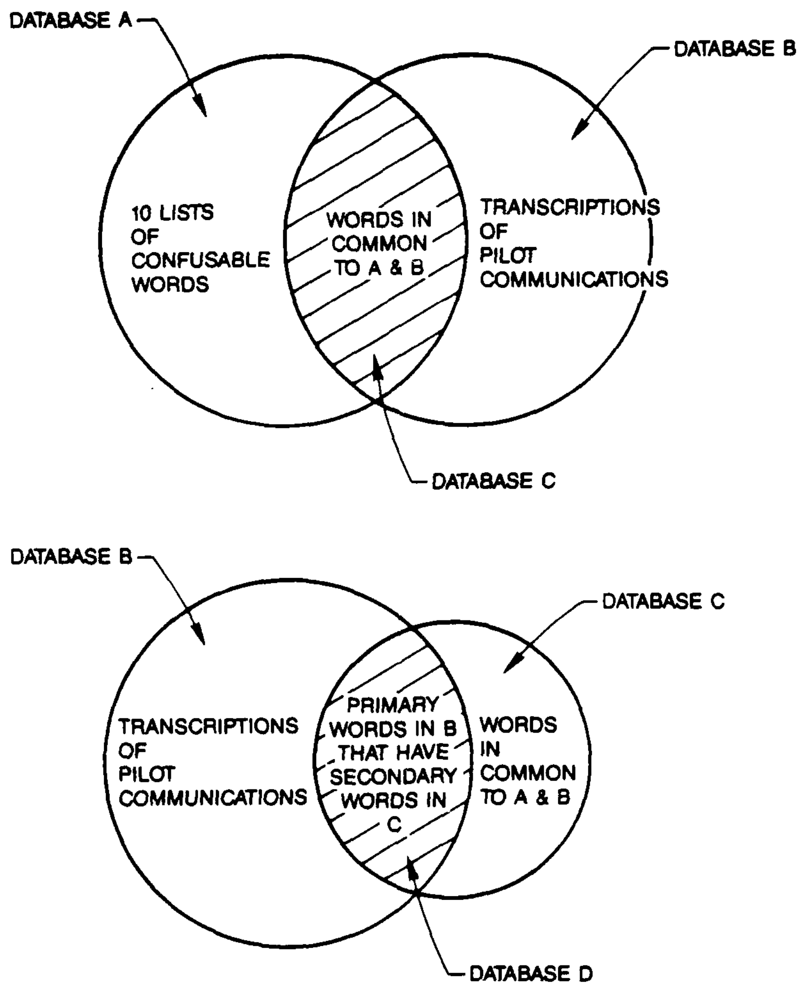

Figure 12. Development of the Scenario Vocabulary

The terms "primary word" and "secondary word" have been used by AAMRL/BBA in research investigations which have included analysis of the information content of specific messages and vocabularies. These terms are simply used to indicate the position each word has or could have within a sentence. A primary word is defined as the first word in a set of two words. Each word in a given vocabulary can be described as a primary word (with the exception of any word which only occurs as the last word in a sentence). Each primary word will have a set of secondary words associated with it. A secondary word is defined as the second word in a set of two words. The set of secondary words following a given primary word will, therefore, consist of any words which could (based upon the linguistic structure of the vocabulary) immediately follow that primary word. For example, consider the following sentences: "Henri Matisse was a great painter," "Jane Austin was a great novelist," "Robert Frost was a great poet." When the word "great" is evaluated as a primary word, the words "painter," "novelist," and "poet" comprise the set of secondary words associated with it.

The actual scenario messages (sentences) were generated by a computer program which utilized the selected scenario vocabulary (database D). This program generated all possible two to six word sentences which followed the linguistic rules defined by database A (i.e., all sentences modeled the natural linguistic structure of the actual pilot communications). All messages were then checked for semantic meaningfulness. Any messages not meeting this criterion were deleted from the set of possible scenario messages. Messages meeting this criterion were randomly combined to form the 30-minute scenario of two to six word sentences. Examples of the actual scenario messages are:

- Two word message: "Turn base."
- Three word message: "My gate nine."
- Four word message: "Change in left turn."
- Five word message: "Winds still at three eight."
- Six word message: "Nine hold wait for flight three."

7.1.2. Construction of the Scenario

The communications scenario has been designed to be a sequence of 40 to 160 messages for each 30-minute time period. Each message will be one of the two- to six-word sentences described in Section 7.1.1. A two-word call sign, individual to each subject, and presented to each subject prior to the start of the scenario, precedes each message. Each call sign presentation occurs in a carrier phrase; for example, "Alpha-One, acknowledge." Immediately after the subject's response to the call sign, the scenario message is presented. A 3- to 5-second break will occur between the end of the subject's response to a message and the presentation of a new message. Each message occurs within the framework of a "timeout period." The timeout period will be the length of time occurring between the end of the message presentation and the time at which the scenario vocabulary disappears from the display. In other words, the timeout period is the length of time the subject has to respond to the message. The timeout period will be visually indicated to the subject by a time clock appearing in the upper right corner of the CRT display. The clock will count down by seconds as the time for message response decreases. When the timeout period is over, the scenario vocabulary will disappear from the display, indicating to the subject that his allotted time for response is over and the time clock will become blank. Table 4 displays the timeout periods associated with the various message lengths.

TABLE 4. TIMEOUT PERIODS FOR VARIOUS MESSAGE LENGTHS

Message Length	Timeout Period
2 words	10 seconds
3 words	15 seconds
4 words	20 seconds
5 words	25 seconds
6 words	30 seconds

7.1.3. Scenario Presentation

Each message in the scenario will be presented verbally to the subject over one of five PACRAT test facility intercom channels (addresses). The subject will hear the message over a set of headphones. The message will be a complete sentence, two to six words in length, which could actually occur in an operational, flight scenario. A time line depicting the message presentation is shown in Figure 13. As soon as the message has been presented, a number of words will be displayed on the three small CRT screens of the subject's test station. If, for example, a six word message was presented, 24 words in six columns of four words each would appear. Two columns of words will appear on each screen. These words will be a partial set of the entire scenario vocabulary (database D). All words in any given column will belong to the same "family" of confusable words (see Section 7.1.1 and Table 3). Each word which occurred in the message will appear in one of the columns. The order in which the words are displayed will be random. The subject's task is to manually select from the CRT screen, within an allotted time period (see Table 4), each of the words which occurred in the message. Each word must be selected in the order that it was presented in the message. Selections are made by pressing the pushbuttons to the left or right of each CRT screen. Figure 14 summarizes the experimenter/subject activities during a scenario message presentation. To select words from columns one, three, and five, the buttons directly to the left of each word will be used. To select words from columns two, four, and six, the buttons directly to the right of each word will be used. The pushbuttons appearing along the top and bottom edges of each display will not be used for word selections, and along with any unused pushbuttons on the six side columns (i.e., not adjacent to a word) will be available for other functions when needed (e.g., "CLEAR," "ENTER," or "RULES"). As each word is selected, it is highlighted and moved, along with its entire column, to the column corresponding to the order in which it was selected. The column of words previously holding that position then moves to the position just

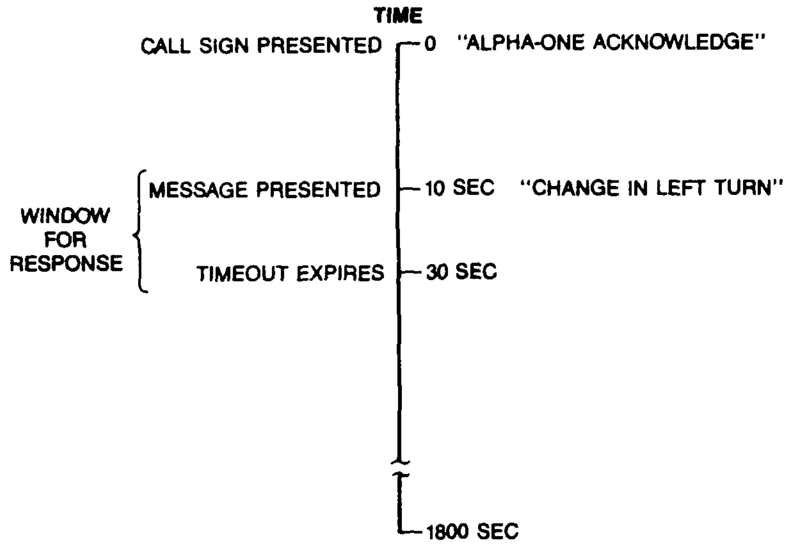

Figure 13. Time Line Depicting the Presentation of the Scenario Messages

Time (Seconds)	Experimenter (E)	Subject (S)	Scenario/Computer
0			Generates Call Sign Message to E
	Transmits Call Sign Message		
		Responds to Call Sign	
			Generates Word Groups on S's CRT
			Generates Communication Message on E's CRT
10	Transmits Communication Message		
		Determines Message Content	
	Retransmits Message	Request Repeats as Required	
20-40*		Performs Message Selection of Repeats, Time Out)	Collects Data (Response Time, Errors, Number)

*Time allowed after message transmisison depends on message length.

Figure 14. Experimenter/Subject Activities During Scenario Message Presentation

vacated by the column containing the selected word. The word selected and subsequent screen change is shown in Figures 15 and 16. For example, the message presented might be "Nine hold wait for flight three." The word "nine" appears in the third column of words on the CRT displays along with the words "none," "night," and "nice." The first column of words includes "wait," "eight," "late," and "rate." As "nine" is selected, it moves along with the other words in its column to column one. The words initially in column one ("wait," "eight," "late," and "rate") then move to column three.

If the subject believes an error has been made, either in the word selected or in the order of the selection, the subject may reselect the entire sentence (provided the timeout period for that particular message has not expired). To reselect a sentence the subject must press the button (one of the available pushbuttons) labeled "CLEAR," and reenter the choice. The subject may also at any time ask for the message to be repeated. When the subject believes the message has been correctly selected, the button labeled "ENTER" will then be pressed. This will input the data for that trial (correct or incorrect response, number of repeats, timeout expired, etc.) into the computer. As soon as a subject has selected the "ENTER" function, or the timeout period has expired, a short (3 to 5 seconds) break will occur. The CRT displays will be blank during this time. At the end of the break a new message will be presented.

7.1.3.1. Linguistic Rules as a Communication Aid

As described in Section 7.1.1, the vocabulary of the scenario was developed to pattern the true linguistic structure of Air Force pilot communication. Like the English language, underlying rules and relationships, both syntactical and semantic, determine the structure of pilot communication. These rules may or may not be absolute. Similarly, these rules may be consciously or unconsciously known to the pilot. An example from the English language might be the knowledge of English speakers that an article would not be followed by a verb. "The" would not be followed by "ran." Adjectives, however, often follow articles. The words "big" and "bird" might often follow "the." These linguistic rules help speakers of English to structure their speech. The English language is, however, made up of a vast vocabulary and many, many rules. For this reason, determining the probability of occurrence of a given word is often difficult. Situational and lexical context often serve as cues, but, again due to the large size of the English vocabulary, this still may be a difficult task.

Fighter pilots, due to their much smaller operational vocabulary, and its more strictly limited situational and lexical contexts of use, may more directly use linguistic rules as communication aids. For this reason, the nonoperational subjects in this research effort will be provided with the

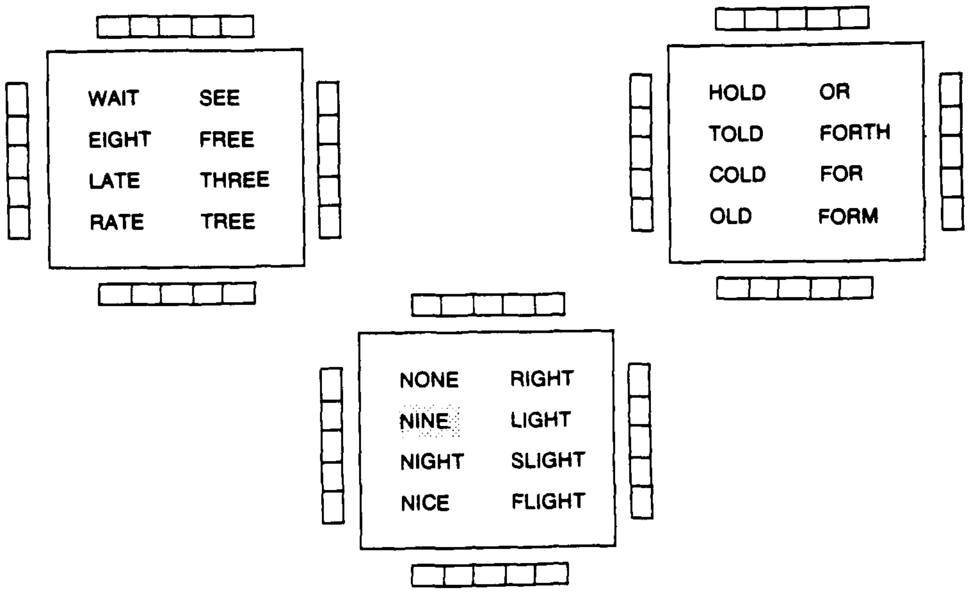

Figure 15. Display Screen as the Subject First Selects the Word "Nine"

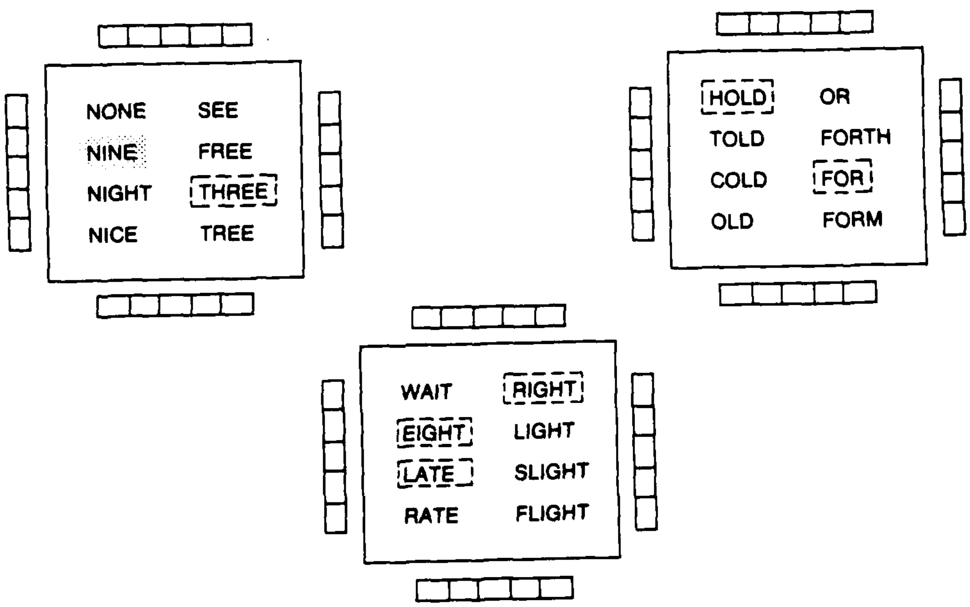

Figure 16. Display Screen Immediately After the Subject's First Selection

linguistic rules that structure the messages of the communication scenario. Although this, of course, will not completely mimic the knowledge and skill that actual pilot subjects will have, it should help to provide a more accurate representation of that knowledge.

Again, the rules provided in the scenario model the linguistic rules found in actual pilot communication (see Section 7.1.1). These rules may be accessed by the subjects at any time during a message presentation. To access the rules for a particular message, the subject must first make a word selection. At that time, the subject may receive a listing of all possible words in the vocabulary which could precede or follow the selected word. This listing will be displayed on the three small CRT screens when the subject presses the button labeled "RULES." Figure 16 shows the selected word (primary) shaded, and all related secondary words on the screen are outlined. The subject should then mark one of the displayed words as the next selection (by pressing the corresponding button). At the discretion of the experimenter, the scenario software may be used with or without the "RULES" function. The experimenter may also wish to teach the subjects the vocabulary and the rule sets in several training sessions prior to the actual testing sessions.

7.1.3.2. HUD Display

The software being implemented for the communication scenario will also include a display of the information given in a standard heads-up display (HUD). Figure 17 illustrates this display. The HUD display will appear along the outer edges of the large CRT screen. Information on this display will include the heading, altitude, and airspeed of the aircraft. This information will be updated throughout the scenario, but will remain constant for the duration of each separate message presentation (updates will occur between message segments). The HUD display may be used with the scenario at the experimenter's discretion. If used, subjects may be requested to verbally report specific aircraft status information (i.e., "State your present altitude"). Requests for status reports should be treated as separate message segments of the scenario. When the scenario is used alone, this display may help to add interest to the task. When the scenario is used in conjunction with a secondary task, this display will serve to direct the subject's attention to the secondary task display (see Section 7.2).

7.2. SECONDARY TASK

Section 4.3 described the utility of presenting secondary tasks concurrently with the task of interest when investigating situations where an operator's performance may be, somehow, degraded. SRL has chosen to provide a secondary tracking task to be used in conjunction with the primary communication task. The tracking task was chosen for a number of reasons. First, tracking tasks

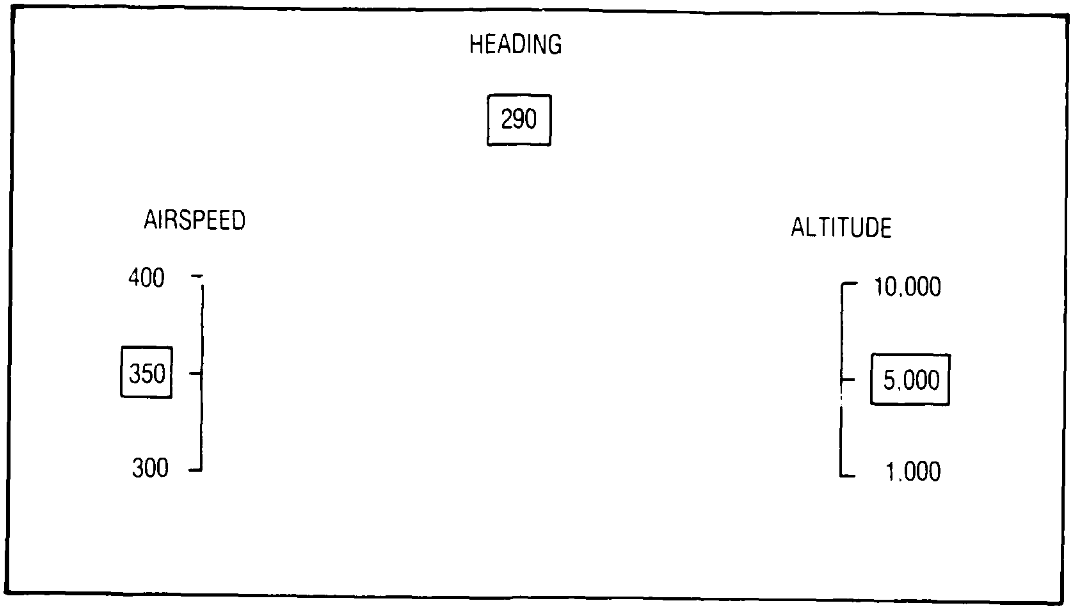

Figure 17. HUD Display

realistically model the flight task. A large portion of a pilot's time is spent either tracking a target or stabilizing a system. Second, in the flight environment tracking tasks and communication tasks will naturally occur together. A pilot is constantly communicating as he flies his aircraft. Third, a large amount of data has been collected using tracking tasks. The reliability, validity, and sensitivity of this data supports the use of tracking as a secondary task. Finally, due to the information processing requirements (Figure 18) of the tracking task, it should not significantly interfere with the communication task. Tracking generally requires a great amount of both visual information processing (visual input, spatial encoding/central processing) and motor output. The communication task described earlier in this section should, instead, require a great amount of auditory processing (auditory input, verbal encoding/central processing) and a minimal to moderate combination of manual and verbal responses.

The tracking task being implemented is a compensatory tracking task of moderate difficulty. The tracking task display will appear in the center of the large CRT screen (Figure 19). The HUD display will appear around the periphery of the tracking display (the dotted lines in Figure 19 would not actually appear, but merely indicate the perimeter in which the secondary task could be presented). Use of the HUD display in the communication scenario will aid in directing the subject's attention to the tracking task. Subject instructions will follow the subsidiary task paradigm (see Section 4.2.2.3.2). Subjects will be told to maintain performance on the communication task at the expense of the tracking task. Separate baseline levels of performance should be collected for both the communication task and the tracking task if both tasks are to be used together. Critical lags

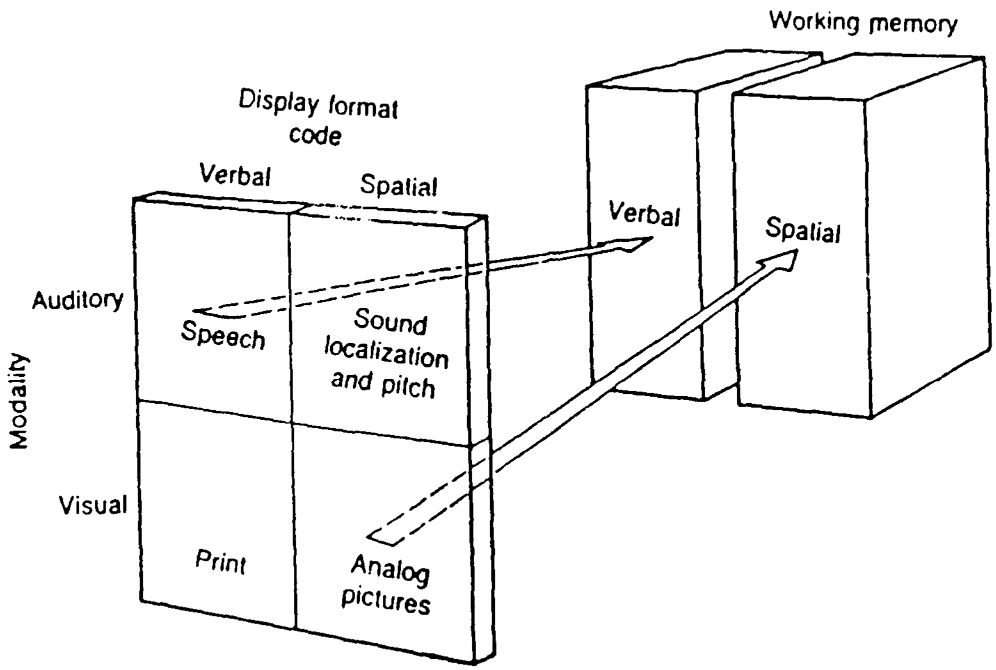

Figure 18. Major Modalities and Display Formats for the Primary Communication Task and the Secondary Tracking Task (Wickens, 1984a)

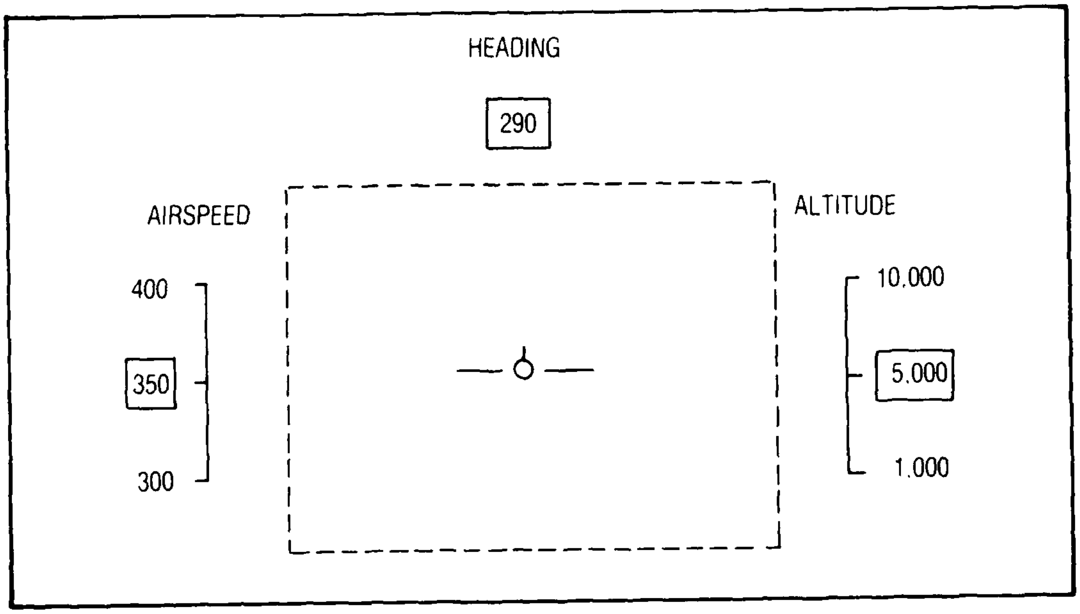

Figure 19. Tracking Task Display

(delays) in the tracking task will be minimal as the joystick system in the PACRAT test facility is force style. Wickens (1988) has shown that with force style sticks where no position feedback is given to the operator, lags in the tracking system can be especially damaging to performance. The use of the tracking task as a secondary task is recommended for AAMRL/BBA's current research interests. However, other secondary tasks (see Section 5.1.2), for example, the linguistic processing task, could easily be implemented on the current system. The remainder of this section will describe the basic components of manual control tasks (tracking tasks) as they relate to the human operator.

7.2.1. The Human Operator as an Element of a Control System

Wickens (1984a) describes a manual control task or "tracking task" as any task in which the control of a dynamic system is accomplished by manipulation of the hands. Manual control tasks differ greatly in their difficulty and demand characteristics, depending upon the system which is to be controlled. Examples of these tasks include driving an automobile, stabilizing an aircraft, or manually assembling miniature components under a microscope. What these tasks have in common is that the operator must continually adjust some control variable to make it correspond to a continuous reference signal. A task may be either to stabilize a system in the face of disturbances (for example, a pilot flying in high winds), or to pursue an evasive target (for example, the target aiming task of a gunner).

A typical tracking task is illustrated in Figure 20 (Wickens, 1986). In this representation a time varying command input, $i_c(t)$, is displayed visually, (D), to the operator, (H). The operator applies some amount of force over time, f(t), to the control device, (C). The resulting control movement, x(t), delivers a signal to the system, (G), which leads to the system response, u(t). The operator exerts control to make u(t) correspond with the command input, $i_c(t)$. This is achieved by minimizing the error, e(t), or the difference between $i_c(t)$ and u(t). Depending upon its characteristics, the display may be described either as "pursuit" or "compensatory." If both $i_c(t)$ and u(t) are presented, the display is a pursuit display. If only their difference, e(t), is presented, the display is described as compensatory. Finally, disturbance inputs, $i_d(t)$, may affect the system. An example of a disturbance input is a gust of wind which moves an aircraft from its approach path. Both disturbances and command inputs represent information presented to the operator. However, disturbance inputs must be corrected, while command inputs are to be followed.

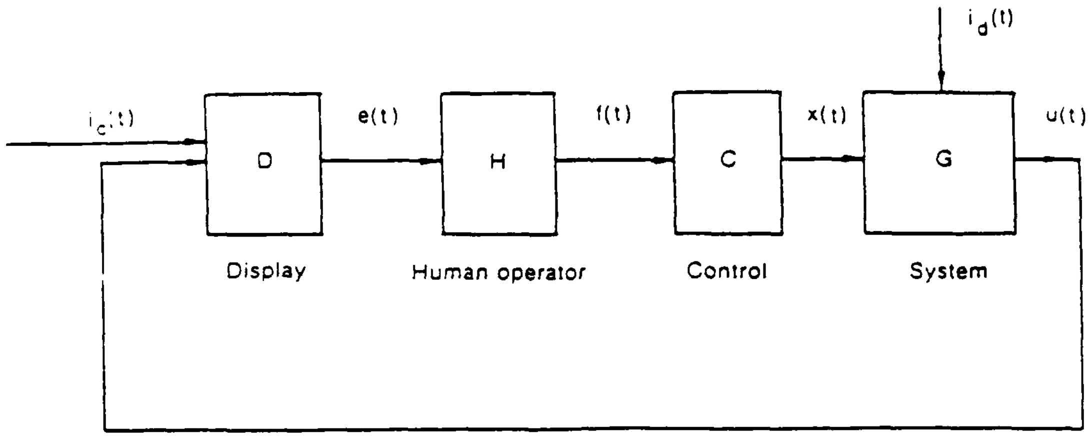

Figure 20. Representation of the Tracking Loop (Wickens, 1986)

7.3. DEPENDENT VARIABLES

Data will be collected on a number of VCE dependent variables. For the communication scenario, the following measures will be collected by the computer:

Time Out	- Subject makes no response in allotted time.
Error	- Subject makes an incorrect verbal or manual response.
Response Time	- Time from the end of a given message transmission by the experimenter to the end of a subject's response.
Repeats	- Number of times the subject asks for a message to be repeated.

For the secondary task, the tracking task, the following measures will be collected.

$$\text{Average Absolute Error (AAE)} = \frac{1}{T}\int_{o}^{T} |e(t)|\, dt$$

$$\text{Root-Mean-Square Error (RMS)} = \sqrt{\frac{1}{T}\int_{o}^{T} e^{2}(t)\, dt}$$

where:

$$T = \text{Task Length}$$
$$e = \text{Error at Time T}$$
$$dt = \text{Sampling Interval}$$

AAE describes the operator's tracking variability when integrated in conjunction with the other performance measures, and RMS is a measurement of performance variability.

7.4. ALTERNATE VCE SCENARIO CONFIGURATION

In addition to the communication scenario described in Section 7.1, an alternate scenario was proposed by SRL but not chosen for implementation. This scenario was based on a menu selection approach modeling the menu selection systems found in some advanced aircraft. The same scenario constraints (see Figure 10) employed during the communication scenario development also served as guidance for the menu selection task. A variety of menu modules were constructed, each module representing a different aircraft system (i.e., WEAPONS). A series of menu pages were constructed for each module, each page representing a different level of that module. Subjects would be required to make a variety of menu selections based upon verbal instructions from the experimenter. Such a scenario could also be easily incorporated with a variety of secondary tasks.

7.4.1. Menu Selection Scenario Design

The design of the menu selection task centered on the existing PACRAT equipment configuration in each of the subject stations (see Section 1). The three small CRTs would be allocated for the presentation of the menus. Subject responses to a menu selection request would be accomplished using the external pushbuttons arranged along the outside edge of the CRTs. These buttons could be considered multifunction keyboards (MFKs).

The main menu (representing various aircraft subsystems, i.e., stores) would be located on the left most CRT. The main menu was comprised of the following aircraft subsystems: communication, navigation, sensors, stores, and systems. Within each of these subsystems, several sublevels were developed to complete the overall menu tree. For example, all of the subsystems were developed to the second sublevel, but only the communication subsystem was developed into a third sublevel. The partial content for the menu tree is depicted in Figure 21.

Main Menu (Aircraft Subsystem)	Communication	Navigation	Sensors	Stores	Systems
Subsystem Level One	UHF VHF IFF	TACAN Flight Mode Doppler/ILS	Radar FLIR ECM	Guns Bombs Missiles	Fuel Landing Gear Lights
Subsystem Level Two	Developed for All Subsystem Level One Items				
Subsystem Level Three	Developed Only for Communication Subsystem Level One Items				

Figure 21. Partial Menu Selection Task Content

Selection of a subsystem would be accomplished by pressing the appropriate MKF key. The middle screen would then display the next sublevel for that subsystem. Activating the MKF key appropriate to the first sublevel would produce a new middle screen depicting the next lower sublevel. This type of menu selection logic is referred to as branching logic. Although the confusability of the menu words is low, in most cases the menu content could be structured with inter or intra sublevel confusability and, thus, obtain greater face validity for the word intelligibility aspects of this scenario.

7.4.2. Menu Selection Scenario Presentation

The presentation of the menu selection task would occur in the same manner as the communication scenario described earlier. The experimenter would, following the scenario script, transmit the instruction to the subject and repeat the message as requested by the subject. Activities required by the subject include perceiving the message and responding to the message through the use of the MFKs. The outcome of the subject's activities would be measures of time outs, errors, response time, and the number of repeats. Figure 22 illustrates the interaction of the experimenter and the subject for an example message presentation during the scenario.

The menu selection scenario could be accomplished independently of other tasks or combined with a secondary task. A tracking task performed in concert with the menu selection activity would provide greater operational construct to the scenario. Other secondary tasks to be used with the MFK menu primary tasks could also be viable for VCE research.

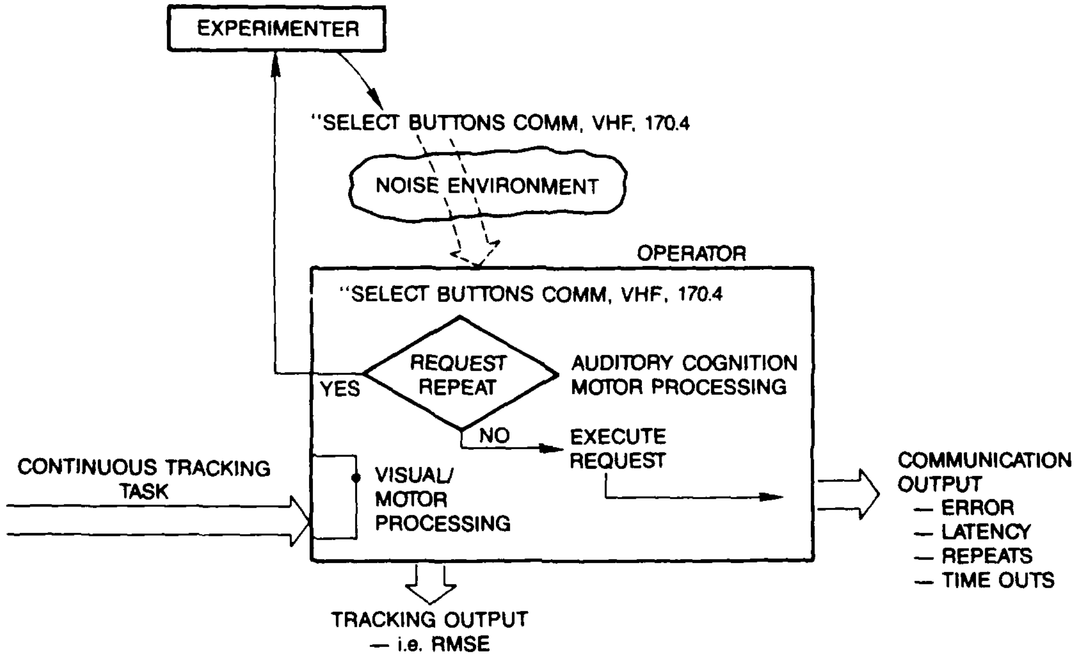

Figure 22. Interaction of the Experimenter and the Subject During a Message Presentation

Section 8
CONCLUSIONS AND RECOMMENDATIONS

8.1. CONCLUSIONS

The evaluation of VCE is a complex problem involving a large number of factors. Research on VCE should include consideration of the following classes of variables:

- Human Component: Information processing requirements and capacities; word recognition, and sentence processing capability in a two-way interactive mode; mental workload.
- Information Component: Syllables, words, sentences, continuous discourse; entropy, mutual information, and channel capacity measurements; task information requirements.
- Equipment Component: Microphones, amplifiers, earphones, jammers, displays; natural/synthetic speech.
- Environmental Component: Noise, acceleration, vibration, physiological stressors (heat/cold, fatigue, pretreatment drugs).

Few actual measures of VCE are reported in the literature, and little theoretical work on the subject of VCE has been discovered. The majority of research on communication effectiveness has traditionally centered on unidirectional rather than interactive two-way communications. Additionally, the literature survey discovered no research on dual-task studies for VCE, although dual-task studies on speech intelligibility were reported.

8.2. RECOMMENDATIONS

Based upon the literature search, the communications scenario development, and the selection of the secondary task, a number of recommendations have been made concerning the evaluation of the selected VCE measure and the expansion of the intelligibility research incorporating the VCE measure.

8.2.1. Evaluation of the Communication Scenario

Because the communication scenario proposed for use in the PACRAT test facility is a newly developed experimental task rather than a standardized psychometric test, no specific validity, reliability, and sensitivity data are available. To ensure generalizability and aid in interpretation of data collected using the VCE scenario, SRL recommends that validity, reliability, and sensitivity evaluations be made.

Validity evaluations (i.e., the extent to which the scenario actually measures the effectiveness of pilot voice communication) should be made by comparing data obtained by using the scenario with data obtained using other standardized intelligibility tests (i.e., MRT, DRT, etc.). Data available for these standardized tests should be positively correlated with data collected using the communication scenario. Validity evaluations should be made for a variety of situations and stressors (i.e., the effect of noise, jamming, workload, etc. on communication effectiveness).

Reliability evaluations (i.e., the consistency, repeatability, or extent to which two applications of the same measure yield the same results) of the communication scenario should also be made. Data collected repeatedly via the scenario under the same experimental conditions should not vary significantly. Reliability evaluations, like validity evaluations, should be made under a variety of experimental conditions.

The sensitivity (i.e., specificity, capability of making fine distinctions) of the communication scenario should also be evaluated. Does data collected using the scenario allow the researcher to assess the relative potential for communication degradations among various equipment design options, various operating conditions, and various task situations? Sensitivity data, like validity data, should be obtained by comparing data collected with the communication scenario to data collected using a variety of other intelligibility measures.

Additional evaluation of the scenario should include the assessment of the number of each message length that constitutes a scenario of low, medium, or high load level for the operator. A primary task with varying levels of load will allow for optimal flexibility of VCE assessment.

8.2.2. Development of Alternative Secondary Task

The use of secondary tasks other than the tracking task will allow for the assessment of other human information processing resources. The tracking task assesses spatial/central processing and a manual response mode. Additional secondary tasks will allow for the examination of processing capacities in conjunction with the primary task (the communication task). For example, a linguistic processing task, although utilizing similar information processing resources as does the primary communication task, would allow for the assessment of performance degradation caused by two competing tasks. Other types of tasks that might prove useful as secondary tasks include tasks requiring information integration, information manipulation, detection, identification, and divided attention. Documentation for a secondary task battery should include the following information for each task: purpose, description, background, reliability, validity, sensitivity, data output specifications, training requirements, and subject instructions.

8.2.3. Additional Research on VCE

Existing VCE related research should be integrated into a theoretical model. Studies should be considered on two-way interactive voice communications where the performance of the operator is dependent on intelligibility of the message. A large number of controlled laboratory experiments should be performed to investigate the individual factors that comprise the communication problem. After these individual factors have been investigated, a group of less structured laboratory experiments, modeling the natural disorganization of human speech and the large variety of task and environmental variables, which comprise real operational situations, should be performed. Finally, field studies (high fidelity aircraft simulator in-flight testing or ground based systems) should be performed in various operational situations.

Appendix A
COMMUNICATION SCENARIO VOCABULARY

Appendix A contains the vocabulary to be used in the communication scenario. The listing that follows is the output of the vocabulary database, database "D2" (see Section 7.1.1 for a description of how the vocabulary was derived). The 496 unique words that comprise the vocabulary are listed in the far left column under the heading "primary word." A primary word is defined as the first word received in any set of two words, or a_k. Explanations of the remaining column headers are as follows:

"FIRST"	if = 0	The primary word at the left cannot appear as the first word in a sentence.
	if = 1	The primary word at the left can appear as the first word in a sentence.
"LAST"	if = 0	The primary word at the left cannot appear at the end of a sentence.
	if = 1	The primary word at the left can appear at the end of a sentence.
"SECONDARY WORD"		The word received immediately after a_k. All words in this column may occur directly after the primary word to the left. The number at the top of this column indicates the number of secondary words in this column.

PRIMARY WORD	FIRST	LAST	SECONDARY WORD
AFT	0	0	0
AIR	1	1	12
			BALL
			CRAFT
			FIVE
			FORCE
			GOOD
			IS
			ONE
			PLAN
			SIX
			TEN
			THEN
			TWO
ALL	1	1	14
			ALL
			AND
			ARE
			AT
			CLEAR
			FIVE
			FOUR
			HAVE
			IN
			NEXT
			RIGHT
			SIZE
			THREE
			WHEN
AM	0	0	3
			AT
			I
			ONE

AND	1	1	110
			ARE
			ASK
			AT
			BACK
			BASE
			BE
			BLAZE
			BRING
			CAPE
			CHECK
			CLEAR
			CLIMB
			COME
			DID
			DO
			EIGHT
			FEEL
			FIND
			FIVE
			FLIGHT
			FLY
			FOR
			FOUR
			GEAR
			GET
			GO
			GOOD
			GOT
			HAS
			HAVE
			HE
			HE'S
			HEAD
			HERE
			HOLD
			HOT
			I
			I'D
			I'LL
			I'M
			IF
			IS
			IT
			JOIN
			JUST
			KEEP
			KILL
			KNOCK
			LEFT
			LET
			LOAD
			LOCKED

LOUD
LOW
MAKE
MODE
NEED
NET
NIGHT
NINE
NO
ONE
PACKS
PASS
PICK
RIGHT
SAME
SAY
SEE
SEND
SET
SHARE
SIX
SLOW
SO
SOME
SOUTH
SQUAWK
STAND
STATE
STICK
STRIKE
STUFF
SWEEP
SWITCH
TAKE
TALK
TEN
THANK
THAT
THEN
THERE
THEY
THEY'LL
THREE
TOUCH
TRY
TURN
TWO
WE
WE'D
WE'LL
WE'RE
WE'VE
WHERE
WHICH
WORK

			WOULD YOU YOU'LL
ARC	0	0	1
			IS
ARE	1	0	24
			AND ARE AT BACK BASE CLEAR DIM DRY FOUR JOINED LOUD ONE RIGHT SOUTH SWEEP THEY THREE TIGHT TWO WE WE'RE WITH WORTH YOU
ASK	0	0	1
			FOR
AT	1	1	30
			ALL BASE EIGHT FIVE FIX FOUR

HIGH
HOT
LEAST
LEFT
LOCK
LOW
MILE
MILES
MY
NIGHT
NINE
ONE
OR
RIGHT
SAY
SIX
SLIDE
TAC
TEN
THAT
THREE
TIME
TWO
WITH

BACK	1	1	21

AND
ARE
AT
CHECK
DRY
FOUR
HERE
IF
IN
LEFT
LIKE
LONG
ONE
RIGHT
SAY
SWITCH
THERE
THREE
WE'RE
WITH
YOU

BAD	0	0	1

TAC

BALL	0	1	14
			AT CLEAR EIGHT FIVE FOUR I IN OR PLAN PLEASE SIX THREE TWO WHEN
BANK	1	0	0
BASE	1	1	22
			AND AT BASE FIVE FOR GEAR I IN IS LEG LOW NINE ONE RANGE SQUAWK STILL THREE TOUCH TURN TWO WATCH WITH
BE	1	1	20
			ALL AT BACK BASE

			CLEAR
			EIGHT
			FIVE
			FLIGHT
			FOR
			FOUR
			IN
			JUST
			NO
			ONE
			RIGHT
			SOME
			THREE
			TWO
			WITH
			YOU
BEACH	1	1	2
			CRAFT
			WAIT
BEAM	0	1	4
			AT
			RIGHT
			TWO
			WITH
BEAR	1	0	3
			ONE
			SIX
			THREE
BEND	0	1	0
BEST	0	0	1
			THING
BET	0	1	0

BIRD	1	0	1
			IT'S
BIT	0	1	5
			IF IN MORE THREE YOU'LL
BLAZE	1	0	2
			EIGHT TWO
BLOCK	0	1	1
			IN
BLOW	0	0	1
			WHY
BLOWN	0	0	0
BLUE	1	0	3
			SIX SOURCE TAIL
BOARD	0	1	3
			AND SPOT TEN

BOOK	0	0	0
BOUND	0	1	6
			AT IS TAKE THREE TYPE YOU
BOX	0	0	0
BRING	0	0	3
			AT IT YOU
BUG	1	0	0
BUT	1	0	17
			DID GO HE HE'S I I'LL IS IT IT'S NO NOT ONE THEN THEY WE WE'LL YOU
BUZZ	0	1	0

BY	0	1	14	AND CHECK CROSS FOR LET MY NIGHT ONE PASS SOME THEY THREE TWO WE'LL
CAB	1	0	2	IT'S ONE
CALL	1	1	13	AND BACK FIVE FOR FOUR I IN IS IT ME MY ONE STRIKE
CAME	0	0	2	BY IN
CAMP	1	0	2	AND ONE

CAN	1	1	30

BE
BRING
CHECK
COME
CROSS
DO
FALL
FIND
GET
GO
HEAR
HIT
I
JUST
MAKE
PICK
PLAN
PUT
REACH
RUN
SEE
SET
STAND
START
TAKE
TELL
TURN
WE
WORK
YOU

CAN'T	0	0	2

GET
SEE

CAP	0	0	1

THREE

CAPE	1	0	1

SIX

CAR	0	0	1	LEFT
CARE	1	0	2	FLIGHT TWO
CASE	1	0	4	FOR ONE THREE TWO
CAT	0	1	3	BALL RED THREE
CHANGE	1	1	3	IN ONE TWO
CHECK	1	1	20	AND BACK FIVE FOR FOUR GAP HIS IN LIST LOW ME PLEASE SPEED THAT TOUCH

				TWO WAIT WILL WITH YOU
CHICK	0	0	1	IN
CHILL	0	1	0	
CHIP	0	0	1	TWO
CLEAN	0	1	1	THROUGH
CLEAR	1	1	20	AND BACK FIVE FOR GO I IT LOUD RIGHT SHOT SHOULD SO STATE STRIKE SWITCH TAKE THAT THREE TOUCH TWO

CLIMB	1	1	1 BACK
CLIP	0	0	1 AND
CLOCK	0	0	1 AND
CLOUD	0	0	0
COACH	0	0	1 ONE
COAST	0	1	0
CODE	1	0	3 FOR IS TWO
COLD	0	0	1 WE
COME	1	1	4 AND BACK IN RIGHT

COMES	0	0	0
COULD	1	1	14
			BE BLOW CALL CHECK COME DO FLY HAVE NOT SWITCH TAKE TALK WE YOU
CRAFT	0	0	1
			DO
CRASH	0	0	1
			IN
CREW	0	1	2
			AND FOR
CROSS	0	1	4
			AT FIVE TWO WIND
CUT	1	0	5

			FOUR GOOD IT PLEASE WE'RE
DARK	0	0	2 RIGHT SPOT
DASH	0	0	1 FIVE
DAY	0	1	0
DECK	0	1	2 TEST WORK
DIAL	1	1	3 FIVE ONE THREE
DID	1	1	5 FIVE IT NOT THEY YOU
DIM	0	1	0

DO	1	1	13
			CHECK DO HAVE IS IT NOT ONE SOME THAT WANT WE WE'D YOU
DOES	1	1	2
			IT NOT
DOG	1	0	1
			FACE
DONE	0	1	0
DOT	0	0	1
			TWO
DROP	1	1	6
			AT FOR IT SIX TANK TWO
DRY	1	1	1
			ONE

```
DUE             0     0          1

                                WEST

DUMP            0     0          1

                                THREE

DUMPED          1     0          2

                                FOUR
                                SIX

EAST            1     1          9

                                BOUND
                                BUT
                                DROP
                                HERE
                                IN
                                SOUTH
                                STATE
                                STRAIGHT
                                SWEEP

EIGHT           1     1         50

                                AND
                                BASE
                                CHANGE
                                CHECK
                                CLEAR
                                EIGHT
                                FIVE
                                FLY
                                FOR
                                FOUR
                                GO
                                HOLD
                                I
                                I'M
                                IF
                                IN
                                IS
                                LEFT
                                LOW
                                MARK
```

				MILE
				MILES
				NINE
				NO
				ONE
				OR
				PLANE
				PLEASE
				RIGHT
				SAY
				SEATS
				SEEN
				SIX
				SPOT
				STEER
				STRIKE
				TELL
				TEN
				TEST
				THANK
				THREE
				TOUCH
				TURN
				TWO
				WE
				WHEN
				WHICH
				WIND
				YOU
				YOU'LL
END	0	1	6	
				FOR
				IF
				IS
				LEFT
				LOW
				PLEASE
EYE	0	1	3	
				BALL
				FLY
				SIX
FACE	0	0	0	

FALL	0	0	1
			IN
FAR	0	1	3
			END LEFT YOU
FAST	0	1	1
			WAY
FEEL	0	0	0
FEET	0	1	7
			AND FOR HE'S IS OR SHOULD TWO
FEW	0	1	0
FIELD	0	1	7
			AT EAST FOR IN IS ONE RIGHT
FILE	0	1	0
FILL	0	1	0

FIND	0	0	1
			YOU
FIRE	1	1	3
			EIGHT WALL WHEN
FIVE	1	1	37
			AIR AND ARE AT BASE BE BOX CALL CHECK CLEAR COULD CREW CROSS DO DUMPED EIGHT FEET FILE FIVE FLIGHT FLY FOR FOUR GO GOOD GOT GRADE GULF HAS HE'S HERE HOLD I'M IF IN IS JUST KNOTS LAST LATE

LEFT
LIFT
LIKE
LOCK
LOUD
LOW
MILE
MILES
NINE
NO
ONE
OR
PART
PLEASE
READ
RIGHT
SAY
SEATS
SIX
SO
SPACE
SPEED
STAND
STATE
STAY
STILL
STRIKE
TAC
TAKE
TAPE
TEN
THAT
THREE
THROUGH
TRUE
TRY
TURN
TWO
WE'LL
WE'RE
WE'VE
WILL
WIND
WITH
WOULD
YOU
YOU'VE

FIX	0	1	1
			WITH

FIXED	1	1	3	
				GEAR THREE WITH
FLEET	0	0	2	
				LOCK TAC
FLIGHT	1	1	24	
				AND CALL CASE CHECK DECK EIGHT FIVE FOUR GATE GO IS KILL ONE PLAN POP RIGHT SIX STRIKE TAKE THREE TWO WILL WITH YOU
FLIP	0	1	0	
FLY	1	1	5	
				BY LEFT PATH THREE THROUGH

FOR	1	1	44
			ALL
			CASE
			CHECK
			EIGHT
			EYE
			FIVE
			FIX
			FLIGHT
			FOUR
			FUEL
			HALF
			I
			IS
			LEFT
			LIGHT
			LIKE
			LINE
			LONG
			LOW
			MARK
			ME
			MEN
			MORE
			MY
			NEW
			NEXT
			NINE
			ONE
			QUIT
			RIGHT
			SHOW
			SIX
			SO
			STOP
			TAKE
			TEN
			THAT
			THREE
			TIME
			TOUCH
			TWO
			WAKE
			WHEN
			YOU
FORCE	0	0	1
			BASE
FORM	0	1	0

FORTH	0	1	0
FOUND	1	0	0
FOUR	1	1	77
			AND
			AT
			BALL
			BASE
			BE
			BEAR
			BUT
			CAN
			CAT
			CHECK
			CLEAR
			COME
			DID
			DO
			DOT
			EIGHT
			FIVE
			FLIGHT
			FOR
			FOUR
			GO
			HALF
			HAS
			HAVE
			HAWK
			HOLD
			I
			IF
			IN
			IS
			IT'S
			JUST
			KNOTS
			LATE
			LEFT
			LOCK
			LOUD
			MILE
			MILES
			MODE
			NINE
			ONE
			OR
			PLEASE
			READ
			RED

				RIGHT SAY SEE SHOULD SIX SOUTH STAND START STATE STEER STILL STOP STRIKE SWITCH TAKE TAP TEN THANK THREE TIME TURN TWO WATTS WE WE'VE WHITE WILL WIND YES YOU YOU'LL
FOURTH	0	0	0	
FREE	0	1	0	
FUEL	1	1	8	HERE LEAK LINE SO STATE STILL THREE WOULD

GAP	0	1	0
GAS	0	1	2
			OR
			WHEN
GATE	0	0	4
			IS
			NINE
			TAC
			THREE
GAVE	0	0	1
			YOU
GEAR	1	1	8
			AND
			CHECK
			FOR
			IS
			LOW
			STOP
			THERE
			TOUCH
GET	1	0	19
			ALL
			BACK
			CAB
			CLEAR
			FIVE
			HERE
			HOLD
			IN
			IT
			MORE
			NEW
			ONE
			RID
			SOME
			THAT

			THROUGH
			TWO
			WITH
			YOU
GLAD	0	0	1
			IT
GLIDE	0	0	3
			PASS
			PAST
			PATH
GO	1	1	23
			ALL
			AND
			AT
			BACK
			BY
			CHECK
			FOR
			GEAR
			GREEN
			HALF
			I'LL
			IF
			IN
			LEFT
			LOW
			ONE
			RIGHT
			ROUND
			THAT
			THREE
			TWO
			WITH
			YOU'LL
GOOD	1	1	20
			CALL
			CHECK
			CLIP
			CUT
			DAY

FLIGHT
FOR
FOUR
GO
GOOD
JUST
LUCK
ONE
SHOW
START
STILL
THEN
TIME
WHEN
YOU'VE

GOT	1	1	24

ALL
BACK
EIGHT
FIVE
FLIGHT
FOR
FOUR
GOOD
HERE
IS
IT
LEFT
ME
MY
ONE
SIGHT
SIX
SOME
THAT
THREE
TWO
WITH
WORD
YOU

GRAB	0	0	0
GRADE	0	0	1

MODE

GRAND	0	0	1 I'LL
GREAT	1	1	0
GREEN	0	1	2 FOR LIGHT
GRID	1	1	2 STATE SWITCH
GROSS	0	1	0
GROUP	1	1	3 FOUR I'LL THREE
GUESS	0	1	1 THAT
GUIDE	0	0	1 IN
GULF	0	1	1 EIGHT
HACK	0	1	0

HAD	0	0	3
			HAD IN NO
HALF	1	1	6
			GLIDE LEFT MILE MILES RIGHT TAKE
HAND	0	1	3
			TURN TWO WING
HANG	0	0	3
			LEFT RIGHT SLIGHT
HAS	1	0	4
			GOT LEFT LOST TWO
HAVE	1	1	20
			EIGHT FIVE HOT IT LET MODE MY NINE NO

NOT
ONE
OR
RAIN
RED
RIGHT
THAT
THREE
TWO
WITH
YOU

HAWK	0	1	3
			AT FOUR ONE
HAZE	0	0	0
HE	1	1	9
			CAN COMES GOT HAS IS JUST SET SHOULD WANTS
HE'LL	1	0	1
			BE
HE'S	1	0	8
			AT IN JUST LOW NO NOT RIGHT STILL

HEAD	0	0	3
			BACK BOUND PLEASE
HEAR	0	1	4
			IS ME THAT YOU
HEARD	0	0	1
			ME
HEIGHT	0	1	0
HELP	1	1	1
			ME
HELPS	0	0	1
			SET
HER	0	1	0
HERE	0	1	12
			AND AT FIVE FOR I'LL IF IN PLEASE TAC THEN

				WE YOU
HIGH	1	1	5	AT FIVE KEY LEFT TWO
HIS	1	0	1	BASE
HIT	0	0	1	BASE
HITS	0	0	1	IN
HO	0	1	2	LEFT YOU
HOLD	1	1	8	FIVE FOR ME MY NINE PLEASE WAIT YOU
HOLE	0	0	1	RING

HOME	0	0	1
			CALL
HOP	0	1	0
HOT	0	1	6
			BUT FIVE LAST SIX THREE TWO
HOW'D	1	0	1
			YOU
I	1	1	35
			AM CALL CAN CAN'T COME COULD DO GAVE GET GO GOT GUESS HAD HAVE HEAR HEARD JUST LEFT LOST MADE NEED ONE READ SAID SAY SEE

SHOW
STILL
THINK
THOUGHT
TOOK
WANT
WENT
WILL
WOULD

I'D	1	0	3

GO
LIKE
SAY

I'LL	1	0	26

BE
BRING
CALL
CHECK
GET
GO
HAVE
HOLD
JOIN
KEEP
LEAVE
LOOK
MAKE
MEET
NEED
PICK
PUT
SAY
SEE
SET
SWITCH
TAKE
TALK
TRACK
TRY
WAIT

I'M	1	0	11

AT
BACK
FIVE

GLAD
HIGH
IN
NOT
SIX
STILL
THREE
WITH

ICE 0 1 0

IF 1 1 12

HE
HE'S
I
IT'S
NOT
SO
THAT
THEY
WE
YOU
YOU'D
YOU'LL

IN 1 1 52

AND
AT
BACK
BASE
BOOK
BOUND
CASE
CHECK
FLEET
FLIGHT
FOR
FOUR
HERE
HIGH
HIS
HOT
I
I'LL
IN
IT
JUST
LEFT

LESS
LOW
MY
NINE
ONE
OR
PLACE
PLEASE
READ
REAL
SAY
SIGHT
SITE
SO
STATE
TANK
TEN
THAT
THEN
THERE
THREE
TOUCH
TOW
TURN
TWO
USE
WANT
WHEN
WITH
YOU

IS 1 1 42

AND
AT
BACK
BASE
CLEAR
COLD
EIGHT
FIVE
FLIGHT
FOUR
GEAR
GOOD
HE
HIGH
IN
IT
JUST
LEFT
LOCK
LOUD
LOW

MODE
NINE
NO
NOT
ONE
PLACED
RED
RIGHT
SIX
SOFT
SPEED
STATE
STILL
STRIKE
TAC
TEN
THAT
THREE
TWO
WET
WITH

IT 1 1 38

ALL
AND
AT
BACK
BLOWN
BY
CHECK
COME
COMES
COULD
DID
FIVE
HERE
IF
IN
IS
IT'S
JUST
MAY
NICE
PAST
RIGHT
SEEMED
SEEMS
SET
SHOULD
SIX
SO
SOUNDS
STILL

			STRAIGHT THAT THEN WHEN WIDE WILL WORKED YOU'VE
IT'S	1	0	19 ALL AT CLEAR COME EIGHT FIVE FOUR I IN JUST LOUD NOT ONE PACK RED SIX STILL THREE TWO
ITS	0	0	0
JET	0	1	5 EIGHT FIVE HE'S IN THAT
JOIN	1	0	2 AT IN

JOINED	0	0	1
			WITH
JOY	0	1	1
			BUT
JUST	1	1	26
			AFT
			BRING
			CALL
			CAME
			DROP
			FALL
			FOUND
			GET
			GO
			HOLD
			IN
			JOINED
			KEEP
			KIND
			MAKE
			MOVED
			NEED
			NO
			NOT
			SOUTH
			STAND
			START
			STAY
			TAKE
			TELL
			WENT
KEEP	1	0	4
			IT
			THAT
			TWO
			YOU
KEY	0	1	2
			THERE
			WITH

KILL	0	0	0
KIND	0	0	0
KNOCK	1	0	1
			IT
KNOT	0	1	2
			SAY
			TWO
KNOTS	0	1	5
			AND
			I
			ICE
			SWITCH
			TWO
LACK	0	0	0
LAND	0	1	10
			AND
			AT
			FIVE
			LINE
			MIGHT
			SO
			THOUGH
			TWO
			WET
			WHEN
LANE	0	1	2
			AND
			STILL

LAST	1	1	5	AT GO RUN TWO WAIT
LATE	1	1	5	FIVE FOUR ONE SIX TWO
LAY	0	0	1	IN
LEAD	1	1	2	ONE THREE
LEADS	1	0	0	
LEAK	0	0	2	IN OR
LEAST	0	0	0	
LEAVE	0	1	2	FOURTH HERE

LEFT	1	1	35
			AND
			AT
			BACK
			BASE
			CLEAR
			CROSS
			CUT
			FIVE
			FOR
			FOUR
			GREEN
			HAND
			HERE
			HOT
			I
			IN
			LONG
			MILE
			NEXT
			NINE
			ONE
			OR
			SIDE
			SIX
			STATE
			TEN
			THERE
			THREE
			TOUCH
			TURN
			TWO
			WHEN
			WHERE
			WIND
			WING
LEG	0	0	1
			IS
LESS	0	1	2
			FIVE
			THAN
LET	1	0	2

			ME YOU
LIFT	0	1	2 AT WINDS
LIGHT	0	1	2 IS IT
LIKE	1	1	13 CLEAR GO I'M IT'S ITS LIVE ME ONE RIGHT SIX SOME THAT THREE TWO WE'RE WINDS YOU YOU'LL
LINE	1	1	6 CLEAR FOR IN THEN THREE TOUCH
LINK	0	0	1 WITH

LIST	0	0	2
			ONE THAT
LIVE	0	0	0
LOAD	0	0	2
			AT IN
LOCK	1	1	1
			ALL
LOCKED	0	1	2
			CLEAR FOR
LOG	0	1	0
LONG	1	1	2
			IS IT'S
LOOK	1	1	4
			AT FOR GOOD LIKE
LOOP	0	1	0

LOST	0	0	5
			ALL HIS IT ME MY
LOT	0	0	1
			WE
LOUD	1	0	3
			AND BUT CLEAR
LOW	1	1	9
			AND KEY LOW SPEED STATE STAY TEN THREE TURN
LUCK	0	1	1
			NEXT
MADE	0	0	1
			THAT
MAIN	0	0	2
			IS SKY

MAKE	1	0	7
			IT
			LEFT
			LOW
			ONE
			SPEED
			THAT
			WAVE
MAN	1	1	1
			SIX
MARK	1	1	5
			AT
			FOR
			FOUR
			TIME
			WHITE
MARKS	0	1	0
MAY	0	0	3
			BE
			HAVE
			PATH
ME	0	1	20
			AT
			BACK
			BASE
			CHECK
			FOR
			HERE
			I
			I'D
			IF
			OR
			PLEASE
			PRIOR
			SAME

				TALK THAT TWO USE WILL WITH WORK
MEAN	0	0	1	RIGHT
MEET	0	1	2	FOR YOU
MEN	1	0	2	AND IN
MERGE	0	0	0	
MID	0	1	0	
MIGHT	0	0	4	BE CHECK HAVE NEED
MIKE	1	1	6	CHECK ONE PUSH RIGHT THREE WIND

MILE	0	1	22
			AND CALL CUT FLIGHT FOR HIGH I IS LEFT ONE PAST PRIOR RIGHT SEA SEE SHOULD SIX SPREAD STRAIGHT TEST THEN THREE
MILES	0	1	36
			AND AT CALL CHECK DO EAST EIGHT FIVE FLIGHT FLY FOR I'D I'M IN IS KEEP LEFT LOCK ONE RIGHT SAY SEE SEEMS SIX SOUTH STATE

			STAY STRAIGHT TAKE THEN THREE TURN TWO WE WEST YOU
MIX	0	0	1 IT
MODE	1	0	5 IS ONE SQUAWK THREE TWO
MOON	1	0	1 TWO
MORE	0	1	4 BASE GUIDE LOW TIME
MOVE	0	0	1 IT
MOVED	0	0	0

MUCH	0	1	5 FOR FUEL WE'LL WOULD YOU
MUST	0	0	1 BE
MY	1	1	14 BASE FUEL GATE GREEN GROUP I LAST LEFT ONE PUSH RIGHT SIDE THREE TURN
NAME	0	0	1 IS
NEAR	0	0	2 AIR END
NEED	1	1	7 ALL FOR IF ME MY SOME YOU

NEEDS	0	0	0	
NET	1	1	6	
				AT
				GO
				ONE
				STRIKE
				THAT
				THREE
NEW	1	0	3	
				ONE
				STATE
				WAKE
NEXT	1	1	5	
				FEW
				ONE
				RUN
				TIME
				TOUCH
NICE	0	0	3	
				AND
				DAY
				FLIGHT
NIGHT	1	1	2	
				THREE
				TIME
NINE	1	1	52	
				ALL
				AND
				ARE
				AT
				BASE

CALL
CAN
CHANGE
CHECK
CLEAR
COULD
CROSS
EIGHT
FIVE
FOR
FOUR
HOLD
I'LL
I'M
IF
IS
KNOTS
LATE
LEFT
MILE
MILES
MODE
NEW
NINE
NO
ONE
PICK
PUSH
RED
RIGHT
SAY
SEATS
SIX
SQUAWK
STATE
STEER
TEN
THANK
THREE
TIME
TURN
TWO
WE'LL
WILL
WIND
WOULD
YOU

NO	1	1	13

BLOW
CHANGE

FOR
GOOD
I
I'M
JOY
NEED
NOT
RANGE
READ
WE
WE'RE

NOSE	0	1	1
			AT
NOT	1	1	18
			AND AT CLEAR COME FLY FOR GET HAVE HOLD IF IN MUCH RIGHT SHOW SO TOO WORK YET
OLD	0	0	1
			ONE
ONE	1	1	115
			AND ARE AT BALL BASE BE

BLUE
CALL
CHANGE
CHECK
CLEAR
COME
DID
DO
EAST
EIGHT
FIND
FIVE
FLIGHT
FLY
FOR
FOUR
FREE
GEAR
GET
GLIDE
GO
GOOD
GOT
HALF
HAS
HAVE
HOLD
HOT
I
I'LL
I'M
IF
IN
IS
ITS
KNOT
LATE
LEFT
LIGHT
LIKE
LOAD
LOCK
LOOK
LOUD
LOW
MADE
MARK
MARKS
MILE
MILES
MORE
MOVED
MY
NEEDS
NINE
NO

ONE
PASS
PLAN
PLEASE
PUSH
READ
RED
RIGHT
SAY
SIDE
SIGHT
SITE
SIX
SOUTH
SPARE
SPEED
SQUAWK
STAND
STATE
STEER
STRAIGHT
STRIKE
SWITCH
TAC
TAKE
TEN
TENTH
TEST
THAT
THEN
THESE
THIRD
THREE
THROUGH
TIGHT
TIME
TRIED
TURN
TWO
WE
WE'LL
WE'RE
WE'VE
WERE
WHEN
WHICH
WHO
WILL
WIND
WITH
WOULD
YOU
YOU'LL

OR	1	0	21
			ARE
			AT
			DO
			FIVE
			FOUR
			GO
			LEFT
			LESS
			NOT
			ONE
			RIGHT
			SIX
			SO
			SWITCH
			THAT
			THREE
			TOUCH
			TWO
			WE
			WHEN
			YOU
PACE	0	0	1
			SPEED
PACK	1	0	1
			ONE
PACKS	0	1	0
PAN	1	0	0
PARK	1	1	2
			AND
			THAT
PARKED	0	0	1
			NEXT

PART	0	1	0
PARTS	0	0	0
PASS	1	1	8 CODE FOUR IT LEFT SIX THAT TWO YOU
PAST	0	1	0
PATH	1	1	10 AND BEAR CALL IS ONE RIGHT SET THREE TURN TWO
PHASE	0	1	1 STATE
PICK	0	0	1 IT
PLACE	0	1	2 AND THREE

PLACED	0	1	0
PLAN	1	1	2 AND FIVE
PLANE	0	1	0
PLATE	0	1	2 AND HOLE
PLEASE	1	1	9 AND CALL FOR GET IF PASS STATE SWITCH THREE
POP	0	0	0
PRIOR	0	0	0
PULLED	0	0	0
PUMP	0	1	1 HE

PUSH	0	0	4 AND FOR THREE TIME
PUSHED	0	0	1 IN
PUT	1	0	4 FOUR IT ONE YOU
QUICK	0	0	1 TWO
QUIT	1	0	0
RACE	0	0	1 HERE
RAID	0	0	1 AT
RAIN	0	0	0
RAISE	0	1	0

RAMP	0	1	0
RAN	0	0	1
			AND
RANGE	0	1	0
RAP	0	0	1
			FOR
RATE	0	1	4
			DO EIGHT JUST THREE
RAY	0	1	6
			AT CHECK IS READ THREE WE
REACH	0	0	1
			ME
READ	1	1	6
			BACK LOUD ME THAT TWO YOU

REAL	1	0	2
			CLEAR
			GOOD
REAR	0	1	0
RED	1	1	9
			CAT
			FOR
			HOT
			LIGHT
			ONE
			RED
			THREE
			TWO
			WHITE
RID	0	0	0
RIGHT	1	1	42
			AND
			AT
			BACK
			BASE
			BE
			CALL
			CAN
			CLEAR
			CROSS
			EIGHT
			FILE
			FIVE
			FOR
			FOUR
			HALF
			HAND
			HERE
			I
			IN
			IS
			LEFT
			LOOK
			LOW
			NEXT

			ONE OR PATH RIGHT SIDE SLIDE TEN THAT THERE THREE TOUCH TURN TWO WE'VE WET WHEN WIND WING
RING	0	0	0
ROAD	0	0	1 SO
ROLL	0	1	1 AT
ROLLS	0	1	0
ROOM	0	1	1 ONE
ROUND	0	0	0
ROW	0	0	1 SO

RUN	0	1	2
			WAY YOU
RUSH	0	0	1
			STAND
SAC	0	0	1
			ONE
SAFE	0	1	0
SAID	0	0	4
			THAT WE WE'LL YOU
SAME	1	1	5
			MODE ONE TWO WINDS WITH
SAVE	1	0	2
			THAT TIME
SAW	0	0	1
			IT

SAY 1 1 21

FLIGHT
I
IT
LOW
MADE
MODE
NEXT
ONE
SAY
STATE
THAT
TYPE
WAY
WE
WE'RE
WE'VE
WEST
WHEN
WHY
WINDS
YOU

SEA 0 0 1

VIEW

SEAL 0 0 1

AND

SEATS 0 1 1

FOR

SEE 1 1 8

BACK
HERE
IF
IT'S
ME
NO
WHERE
YOU

SEEM	0	0	1 LIKE
SEEMED	0	0	0
SEEMS	0	0	0
SEEN	0	0	1 LOCK
SEND	1	0	2 HERE IT
SET	1	1	4 IT THESE WOULD YOU
SHACK	0	1	0
SHARE	0	1	0
SHE'S	1	0	1 STILL
SHIP	0	1	1 TAKE

SHOT	0	1	1
			IN
SHOULD	1	0	6
			BE CRASH HAVE KEEP READ SEE
SHOW	0	1	5
			CALL IT ME THREE YOU
SIDE	0	1	3
			AND FOR THERE
SIGHT	0	1	6
			AND AT BUT FOR GEAR TEN
SITE	0	1	2
			DOES PLEASE
SIX	1	1	30

AND
ARE
AT
BACK
BALL
BASE
BE
CALL
CAN
CHANGE
CHECK
CLEAR
DASH
DO
DOES
EIGHT
FIVE
FLIGHT
FLY
FOR
FOUR
GATE
GO
GOOD
GRAB
HAS
HE'S
I
I'D
I'LL
I'M
IF
IN
IS
KNOTS
LAST
LATE
LEFT
LIKE
LOCK
LOUD
LOW
MAKE
MILE
MILES
NINE
NOT
ONE
OR
PACE
RED
RIGHT
SAY
SIX
SPEED

			SQUAWK
			STAND
			STATE
			STAY
			STILL
			STRIKE
			SWEEP
			SWEET
			SWITCH
			TAKE
			TEN
			TEST
			THEN
			THREE
			TIME
			TURN
			TWO
			WE'RE
			WELL
			WILL
			WIND
			WITH
			WOULD
			YOU
			YOU'LL
SIZE	0	0	0
SKIP	0	1	0
SKY	1	0	1
			CLEAR
SLIDE	0	1	2
			STEER
			TWO
SLIGHT	0	0	5
			ICE
			LOT
			RIGHT
			STEER
			TAIL

SLING	0	0	0
SLOW	1	1	0
SO	1	1	11 FAR FORTH I I'M LONG MUCH THAT WE WE'LL YOU YOU'LL
SOFT	1	1	0
SOME	0	1	10 AIR FUEL GAS HAZE I KIND MORE NEAR PARTS SLIGHT
SOON	0	1	1 STAY
SOUND	0	0	1 WE'LL

SOUNDS	0	1	1
			GOOD
SOURCE	0	0	1
			AND
SOUTH	0	1	4
			AT
			BOUND
			FAST
			HERE
SPACE	0	0	0
SPARE	0	0	1
			ONE
SPEED	1	1	10
			AND
			AT
			FIVE
			FOR
			FOUR
			ONE
			SIX
			TEN
			THREE
			TWO
SPLIT	0	0	1
			DO
SPOT	0	0	4

				FIVE FOUR THREE TURN
SPREAD	0	1	1	SLIGHT
SQUAWK	1	1	5	AND FOUR ONE SIX STAND
STAGE	0	0	1	RIGHT
STAND	1	0	4	BY CLEAR IS IT
STAR	1	0	1	ONE
START	1	1	3	AT KIND LIVE
STATE	1	1	20	

				AND BASE CASE EIGHT FIVE FOR FOUR FUEL IS MODE NINE ONE PLEASE RIGHT SIX TEN THREE TWO WE'RE WHEN
STAY	1	0	12	AT CLEAN CLEAR DID FOR HERE IN LEFT SIX TEN WEST WITH
STEER	1	1	0	
STERN	0	1	0	
STICK	0	0	2	IT'S RIGHT

STILL	1	1	8
			CLEAN GOT HAVE NO SIX THREE WITH WORK
STOP	1	1	10
			AND AT CHECK FOR LAST NO OR STRAIGHT THEN WITH
STORE	0	0	2
			BACK YET
STORES	0	0	1
			ARE
STRAIGHT	1	1	4
			BASE IN THROUGH YOU
STRIKE	1	1	32
			AND ARE AT

			BASE
			CALL
			COULD
			DID
			DO
			DOES
			FIVE
			FOUR
			GET
			GO
			HAS
			HE'S
			I
			IF
			IS
			MODE
			NO
			ONE
			SIX
			STRIKE
			SWITCH
			TEN
			TEST
			THERE
			TWO
			WE'LL
			WITH
			YET
			YOU
STUFF	0	1	0
SWEEP	0	1	3
			ARE
			PART
			STAND
SWEET	1	1	0
SWITCH	1	1	8
			AND
			BACK
			IN
			IT

			ONE PATH PLEASE YOU
TAC	1	1	11
			AND FIVE FOUR HERE IN ONE ROOM SIX TEN THREE TWO
TAIL	0	1	8
			AND FOUR RED SIX TAP THREE WIND YOU
TAKE	1	1	13
			EIGHT IN IT LEFT LOW MODE ONE SOME SPEED TEN THAT TWO YOU
TALK	1	0	0

TANK	1	1	3
			FUEL THAT TWO
TANKS	1	0	0
TAP	1	0	3
			ONE ROOM TWO
TAPE	1	0	0
TAPPED	0	1	0
TASK	0	1	0
TELL	1	0	2
			FIVE ME
TEN	1	1	27
			AND ARE DAY DIAL EIGHT FIVE FOUR I'M IS LIGHT MAKE MILE MILES NEED NINE

			NO
			ONE
			RIGHT
			SAY
			SIX
			SPACE
			STATE
			STRIKE
			SWITCH
			TAKE
			TWO
			WOULD
TENTH	0	0	0
TEST	1	1	12
			CHANGE
			DONE
			FLIGHT
			FOUR
			IS
			IT
			LOUD
			ONE
			SIX
			TEST
			THREE
			TWO
THAN	0	0	1
			FIVE
THANK	1	0	1
			YOU
THAT	1	1	47
			ALL
			AND
			ARE
			BALL

BLUE
BUT
CALL
CLEAR
COULD
DID
FIVE
FOR
GOT
I
I'D
I'M
IN
IS
JUST
LAST
LIGHT
LIST
LONG
MIGHT
NEED
NINE
NOT
ONE
RIGHT
SAME
SHOULD
SIX
THAT
THERE
THEY
THING
THREE
TIME
TWO
WAY
WE
WE'RE
WERE
WHEN
WILL
YOU
YOU'VE

THEN	1	1	14

BACK
BE
CHECK
HOLD
I'LL
RIGHT

TURN
TWO
WE
WE'D
WE'LL
WE'RE
YOU
YOU'LL

THERE 1 1 22

AND
AT
FOUR
I
I'LL
I'M
IF
IN
IS
JUST
LIKE
OR
PLACE
PLEASE
PUT
RIGHT
SEEMS
SHOULD
SO
WHEN
WILL
YOU

THESE 0 1 1

IS

THEY 1 0 13

ARE
COME
GOT
HAD
HAVE
MIGHT
MUST
NEED
PUT
SAID

			SAY SHOULD START
THEY'LL	1	0	3
			BE GET HAVE
THEY'VE	0	0	1
			GOT
THICK	0	1	0
THING	0	0	1
			COMES
THINK	0	1	8
			I I'M IT SO THAT THEY TWO YOU
THIRD	1	0	2
			MAN WORD
THOUGH	0	1	0
THOUGHT	0	0	1

IT

THREE	1	1	36

AND
ARE
AT
BACK
BASE
BE
BLUE
BUT
CAN
CHECK
CLEAR
COME
COULD
CREW
CUT
DID
DO
DROP
EIGHT
FIVE
FLIGHT
FLY
FOR
FOUR
GEAR
GO
GULF
HAS
HAVE
HE'S
HOLD
I
I'D
I'LL
IF
IN
IS
IT'S
JUST
KNOTS
LATE
LEFT
LOUD
MAKE
MILE
MILES
NINE
NO
ONE
OR

				REACH
				REAR
				RED
				RIGHT
				ROLLS
				SAY
				SIX
				SLIDE
				SOUNDS
				SOUTH
				SQUAWK
				STAND
				STATE
				STAY
				STILL
				STRIKE
				SWITCH
				TAC
				TAKE
				TEN
				THAT
				THREE
				TOUCH
				TRY
				TURN
				TWO
				USE
				WE
				WE'LL
				WE'RE
				WHEN
				WILL
				WIND
				WINGS
				YOU
				YOU'LL
THROUGH	0	1	6	
				GRID
				IT
				SIX
				SPEED
				TEN
				WITH
THRUST	0	0	2	
				PLATE
				SEAL

TIGHT	0	0	2 AND TURN
TIME	1	1	21 AND ARE AT BUT FIVE FOR FOUR HACK IN IS ONE SIX SO STATE THREE TWO WE WE'RE WHICH WILL YOU
TOLD	0	0	0
TOO	0	1	3 FAR LOW MUCH
TOOK	0	0	1 ONE
TOP	0	1	2 CAP SLING

TOPPED	0	0	0
TOUCH	1	1	2 AND FOUR
TOW	1	1	1 SIX
TRACK	0	0	1 IT
TREE	0	0	1 CAN
TRICK	0	0	1 YOU
TRIED	0	0	0
TRUE	1	1	3 THREE TURN TWO
TRY	1	1	4 AND IT ONE WE'RE

TURN 1 1 19

AND
AT
BACK
BASE
CROSS
END
FOR
HIGH
IN
IT
LEFT
NEXT
ONE
RIGHT
THAT
THERE
TWO
WHEN
YOU

TWO 1 1 92

AIR
AND
ARE
AT
BASE
BE
BEAM
CALL
CAN
CHANGE
CHECK
CLEAR
COME
CREW
CROSS
DID
DO
EAST
EIGHT
FIVE
FLEET
FLY
FOR
FOUR
GO
GOT
HALF
HAS
HIS
HOLD

I
I'LL
I'M
IF
IN
IS
JUST
KNOTS
LATE
LEFT
LIKE
LOCK
LOG
MARK
MILE
MILES
MORE
NINE
NO
ONE
OR
RATE
READ
RED
RIGHT
ROLL
SAFE
SAY
SEE
SIX
SO
SOUND
SOUTH
SPEED
START
STATE
STAY
STEER
STOP
STRAIGHT
STRIKE
SWITCH
TAC
TAKE
TEN
TEST
THANK
THAT
THREE
TRY
TURN
TWO
WE
WE'RE
WEST
WHEN

				WILL WIND WITH YOU YOU'LL YOU'VE
TYPE	1	0	2	AND ONE
USE	1	1	3	BEAM SOME THAT
VEER	1	0	0	
VIEW	0	0	0	
WAIT	1	1	5	AND FOR I ONE THAT
WAKE	1	0	0	
WALL	0	0	0	
WANT	0	1	6	FIVE ME ONE SOME THAT YOU

WANTS	0	0	0
WATCH	1	1	0
WATTS	0	1	0
WAVE	0	0	1
			IN
WAY	0	1	12
			AND
			ARE
			CAN'T
			DO
			FIVE
			FOR
			HE
			SO
			STEER
			THREE
			WE
			YOU
WE	1	1	32
			ARE
			BRING
			CAME
			CAN
			CLIMB
			COME
			COULD
			DO
			FIND
			GET
			GO
			GOT
			HAD
			HAVE
			JUST
			LOOK
			MADE
			NEED
			REACH
			READ

			SAID SAW SEE SHARE SHOW STILL TOLD TRIED WANT WERE WILL WOULD
WE'D	1	0	3
			JUST LIKE TAKE
WE'LL	1	0	28
			BE BRING CALL DO GET GO HAVE JOIN JUST LET LOOK MAKE MEET NEED PUT RUN SEE SET STAND STAY STILL SWITCH TAKE TALK TRY TURN USE WAVE

WE'RE	1	0	11
			AT
			BACK
			CLEAN
			CLEAR
			FOUR
			IN
			JUST
			NOT
			ONE
			THREE
			TOUCH
WE'VE	1	0	4
			GOT
			HAD
			LET
			LOST
WEAK	0	1	2
			AND
			SWITCH
WELL	1	1	9
			IT
			IT'S
			LEFT
			RIGHT
			SHE'S
			TELL
			WE
			WE'VE
			YOU
WENT	0	0	0
WERE	1	0	4
			CUT
			LOUD
			NOT
			YOU

WEST	0	1	10	AND AT BOUND FLY HERE LOW OR PAST SHOULD SIDE
WET	1	1	1	TOO
WHEN	1	0	9	DID IT'S NEAR SPEED THAT THEY WE WOULD YOU
WHERE	1	0	5	ARE I'M IS WE WE'RE
WHICH	1	0	5	FOUR IS SHOULD WAY WILL

WHILE	0	0	3	
				I
				IT'S
				WE'RE
WHITE	0	1	1	
				ONE
WHO	1	1	5	
				GO
				HE
				PULLED
				WANTS
				WILL
WHOLE	0	0	1	
				LOT
WHY	1	1	2	
				I
				IT
WIDE	0	0	4	
				LEFT
				OR
				SO
				THREE
WILL	1	1	16	
				BE
				CHECK
				DO
				FLY
				GO
				HIT
				HOLD

			I KEEP MAKE NEED NOT PUT SET TAKE YOU
WIND	1	1	7
			CHECK FOR HERE IS ONE THREE TWO
WINDS	1	1	10
			ARE AT FOR ONE PLEASE RIGHT STILL THREE TWO WILL
WING	0	1	3
			I RIGHT WE
WINGS	0	0	1
			WE
WITH	1	1	18
			BASE BLUE

			EAST FOUR FUEL GEAR IT ME MY NO ONE SIX SOME THAT THREE TWO WITH YOU
WORD	0	1	0
WORK	1	1	7 AND I'M IN IT SLOW THAT WEST
WORKED	0	0	0
WORTH	0	0	0
WOULD	1	0	5 BE LIKE PLEASE WE'D YOU
YANK	0	0	1 YOU

YES	1	1	2
			CAN
			WE'VE
YET	0	1	3
			AT
			BUT
			DO
YOU	1	1	112
			ALL
			AND
			ARE
			AT
			BACK
			BASE
			BE
			BET
			BRING
			CALL
			CAN
			CHECK
			CLEAR
			CLIMB
			CLOUD
			COME
			COULD
			CUT
			DID
			DO
			DROP
			EIGHT
			FEEL
			FIND
			FIRE
			FIVE
			FLY
			FOR
			FOUND
			GEAR
			GET
			GO
			GOOD
			GOT
			HAD
			HAVE
			HEAD
			HEAR
			HEARD

HERE
HIT
HOLD
I
I'LL
I'M
IF
IN
JOINED
JUST
KEEP
KNOCK
LACK
LEAVE
LIKE
LOOK
LOST
LOUD
MAKE
MARK
MAY
MEAN
MERGE
MIGHT
MOVE
MOVED
MUCH
NEED
NET
NOT
ONE
OR
PASS
PICK
PLEASE
PUT
READ
RIGHT
RUN
SAID
SAME
SAY
SEE
SET
SHOULD
SLIDE
STAND
START
STATE
STAY
STILL
TALK
TELL
THERE
THREE
TOO
TOPPED
TRIED

TURN
TWO
USE
WANT
WATCH
WEAK
WERE
WHEN
WILL
WITH
WORK
WOULD
YET
YOU
YOU'LL

YOU'D	0	0	3

LIKE
PLEASE
PUT

YOU'LL	1	0	3

BE
HAVE
STAY

YOU'VE	1	0	1

GOT

BIBLIOGRAPHY

Aasterud, J. P. (1985). RLPM--A communicator's planning tool. In the Proceedings of MILCOM '85. IEEE Military Communications Conference, 353-358. Boston, Massachusetts: Institute of Electrical and Electronic Engineers.

Acton, W. H., Crabtree, M. S., Simons, J. C., Gomer, F., and Eckel, J. S. (1983). Quantification of crew workload imposed by communications-related tasks in commercial transport aircraft. In the Proceedings of the Human Factors Society 27th Annual Meeting, 239-243. Santa Monica, California: Human Factors Society.

Adams, M. J. (1979). Models of word recognition. Cognitive Psychology, 11, 133-176.

Aldrich, T. B. and Szabo, S. M. (1986). A methodology for predicting crew workload in new weapon systems. In the Proceedings of the Human Factors Society 30th Annual Meeting, 633-637. Santa Monica, California: Human Factors Society.

Allen, J. (1981). Linguistic-based algorithms offer practical text-to-speech systems. Speech Technology, Fall, 12-16.

Anderson, T. R., McKinley, R. L., and Moore, T. J. (1982). Evaluation of speech synthesis for use in military noise. In the Proceedings of National Bureau of Standards Workshop on Standardization, 241-245. Washington, D.C.: National Bureau of Standards.

Anderson, T. R. and Moore, T. J. (1984). Characterizing requirements of speech I/O for military cockpit environments. In the Proceedings of Speech Tech '84, 77-80. New York: Media Dimensions.

Anderson, T. R., Moore, T. J., and McKinley, R. L. (1985). Issues in the development and use of a speech recognition database for military cockpit environments. In the Proceedings of Speech Tech '85, 172-176. New York: Media Dimensions.

Annett, J. and Paterson, L. (1967). Training for auditory detection. In A. F. Sanders (Ed.), Attention and Performance, 420-426. Amsterdam: North-Holland. (Reprinted from Acta Psychologica, 27, 1967.)

ANSI (1977a). American National Standard Criteria for Permissible Ambient Noise During Audiometric Testing, No. ANSI S3.1-1977. New York: American National Standards Institute, Inc.

ANSI (1977b). American National Standard for Rating Noise with Respect to Speech Interference, No. ANSI S3.14-1977. New York: American National Standards Institute, Inc.

Arbak, C. J., Greene, J. H., Shew, R. L., Simons, J. C., and O'Donnell, R. D. (1984). Ground Attack Tactics Survey (GATS): IV. Integrated Analysis of Pilot Workload, Technical Report No. AFAMRL-TR-84-047. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Aretz, A. J. (1984). A comparison of manual and voice response modes for the control of aircraft systems. In the Proceedings of the Human Factors Society 28th Annual Meeting, 97-101. Santa Monica, California: Human Factors Society.

Armstrong, J. W. and Poock, G. K. (1981). Effect of Task Duration on Voice Recognition System Performance. Monterey, California: Naval Postgraduate School.

Baker, J. M. and Pinto, D. F. (1986). Optimal and suboptimal training strategies for automatic speech recognition in noise, and the effects of adaptation on performance. In the ICASSP 86 Proceedings. International Conference on Acoustics, Speech, and Signal Processing, Vol. 1 of 4, 745-748. Tokyo: Institute of Electrical and Electronic Engineers.

Beeson, W. R. (1977). An algorithm for determining speech intelligibility (Thesis), Technical Report No. AFIT/GE/EE/77-9. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, DTIC No. ADA055236.

Berliner, C., Angel, L., and Shearer, D. J. (1964). Behaviors, measures, and instruments for performance evaluation in simulated environments. Paper presented at the Symposium and Workshop on the Quantification of Human Performance. Albuquerque, New Mexico.

Biers, D. W. (1984). Voice Stress as a Workload Measure: Review of the Literature and Recommendation for Further Study. Dayton, Ohio: Systems Research Laboratories, Inc. (Unpublished Report).

Binford, J. R. and Loeb, M. (1966). Changes within and over repeated sessions in criterion and effective sensitivity in an auditory vigilance task. Journal of Experimental Psychology, 72, 339-345.

Boff, K. R., Kaufman, L , and Thomas, J. P. (Eds.). (1986). Handbook of Perception and Performance, Vol. I, Sensory Processes and Perception. New York: Wiley.

Boggs, D. H. and Simon, J. R. (1968). Differential effects of noise on tasks of varying complexity. Journal of Applied Psychology, 52, 148-153.

Bolt, R. F. and Sanderson, B. D. (1981). A multiplexed digital voice in communications system compatible with future voice conversion techniques. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 26.1-26.7. London: Technical Editing and Reproduction Ltd.

Bond, Z. S. and Garnes, S. (1980). Misperceptions of fluent speech. In R.A. Cole (Ed.), Perception and Production of Fluent Speech (Chapter 5). Hillsdale, New Jersey: Lawrence Erlbaum.

Bradshaw, J. L. (1984). A guide to norms, ratings, and lists. Memory and Cognition, 12, 202-206.

Brandenburg, K. (1987). OCF-A new coding algorithm for high quality sound signals. In the Proceedings: ICASSP 87, 1987 International Conference on Acoustics, Speech, and Signal Processing, 141-144. Dallas, Texas: Institute of Electrical and Electronic Engineers.

Braune, R. and Wickens, C. D. (1983). The functional age index: An objective decision criterion for the assessment of pilot performance capabilities. In the Proceedings of the Second Symposium on Aviation Psychology, 253-268. Columbus, Ohio: Department of Aviation Psychology.

Braune, R. and Wickens, C. D. (1986). Time-sharing revisited: Test of a componential model for the assessment of individual differences. Ergonomics, 29, 1399-1414.

Broadbent, D. E. (1957). Effects of noise on behavior. In C. M. Harris (Ed.), Handbook of Noise Control. New York: McGraw-Hill.

Broadbent, D. E. (1978). The current state of noise research: Reply to Poulton. Psychological Bulletin, 85, 1052-1067.

Broadbent, D. E. (1982). Task combination and selective intake of information. Acta Psychologica, 50, 253-290.

Broadbent, D. E. and Gregory, M. (1964). Accuracy of recognition for speech presented to the right and left ears. Quarterly Journal of Experimental Psychology, 16, 359-360.

Brown, I. D. (1966). Subjective and objective comparisons of successful and unsucessful trainee drivers. Ergonomics, 9, 49-56.

Brown, I. D. (1978). Dual task methods of assessing workload. Ergonomics, 21, 221-224.

Bryden, M. P. (1963). Ear preference in auditory perception. Journal of Experimental Psychology, 65, 103-105.

Burch, W. M. and Butlin, A. T. (1980). A microcomputer-based simultaneous vigilance test. Perceptual and Motor Skills, 50, 1195-1202.

Butterbaugh, L. C. (1978). Crew Workload Technology Review and Problem Assessment, Technical Memorandum No. AFFDL-TM-78-74-FGR. Wright-Patterson Air Force Base, Ohio: Flight Dynamics Laboratory.

Carhart, R. Tillman, T. W., and Greetis, E. (1969). Perceptual masking in multiple sound backgrounds. The Journal of the Acoustical Society of America, 45, 694-703.

Carswell, C. M. and Wickens, C. D. (1985). Lateral task segregation and the task-hemispheric integrity effect. Human Factors, 27, 695-700.

Casali, J. G. and Wierville, W. W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. Ergonomics, 27, 1033-1050.

Chapanis, A., Garner, W. R., and Morgan, C. T. (1949). Applied Experimental Psychology, Human Factors in Engineering Design (Chapter 8, Speech). New York: Wiley.

Cherry, E. C. and Taylor, W. K. (1954). Some further experiments upon the recognition of speech, with one and with two ears. The Journal of the Acoustical Society of America, 26, 554-559.

Childress, M. E., Hart, S. G., and Bortolussi, M. R. (1982). The reliability and validity of flight task workload ratings. In the Proceedings of the Human Factors Society 26th Annual Meeting, 527-532. Santa Monica, California: Human Factors Society.

Chiles, W. D. and Alluisi, E. A. (1979). On the specification of operator or occupational workload with performance measurement methods. Human Factors, 21, 515-528.

Clement, W.F., Jex, H.R., and Graham, D. (1968). A manual control-display theory applied to instrument landings of a jet transport. IEEE Transactions on Man-Machine System, MMS-9, 93-110.

Cliff, R. C. (1973). Attention sharing in the performance of a dynamic dual task. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3, 241-248.

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. Perception and Psychophysics, 1, 153-156.

Connor, S. A. (1981). A Comparison of Pilot Workload Assessment Techniques Using a Psychomotor Task in a Moving-Base Aircraft Simulator. Master's Thesis, Virginia Polytechnic Institute and State University, Blackburg, Virginia.

Costello, R. G. (1976). Continuous compensatory audio manual tracking. In the Proceedings of the Twelfth Annual Conference on Manual Control, Technical Memorandum No. NASA-TM-X-73. Washington, D.C.: National Aeronautics and Space Administration.

Crawford, B. M., Pearson, W. H., and Hoffman, M. S. (1978). Multipurpose Switching and Flight Control Workload, Technical Report No. AMRL-TR-78-43. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory.

Crosby, J. V. and Parkinson, S. R. (1979). A dual task investigation of pilots' skill level. Ergonomics, 22, 1301-1313.

Damos, D. L. (1978). Residual attention as a predictor of pilot performance. Human Factors, 20, 435-440.

Damos, D. L. (1985). The effect of asymmetric transfer and speech technology on dual-task performance. Human Factors, 27, 409-421.

Damos, D. L. (1986a). The effect of using voice generation and recognition systems on the performance of dual tasks. Ergonomics, 29, 1359-1370.

Damos, D. L. (1986b). The role of asymmetric transfer in the evaluation of voice generation and recognition systems. In the AGARD Conference Proceedings - Information Management and Decision Making in Advanced Airborne Weapon Systems. London: Technical Editing and Reproduction Ltd.

Damos, D. L. (1986c). The Use of Interactive Voice Systems to Improve Multiple-Task Performance, Report No. 916416-FTN. Arlington, Virginia: Office of Naval Research, DTIC No. ADA171492.

Damos, D. L., Bittner, A. C., Jr., Kennedy, R. S., and Harbeson, M. M. (1981). Effects of extended practice on dual-task tracking performance. Human Factors, 23, 627-731.

Damos, D. L., and Smist, T. E. (1983). Individual differences in multiple-task performance as a function of response strategy. Human Factors, 25, 215-226.

Department of Defense. (1981). Military Standard - Human Engineering Design Criteria for Military Systems, Equipment, and Facilities, MIL-STD-1472C. Washington, D.C.: U.S. Government Printing Office.

Deutsch, D. (1986). Auditory pattern recognition. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance, (Chapter 32). New York: Wiley.

Doll, T. J., and Folds, D. J. (1985). Auditory signals in military aircraft: Ergonomic principles versus practice. In the Proceedings of the IEEE 1985 National Aerospace and Electronics Conferece, 958-965. Dayton, Ohio: Institute of Electrical and Electronic Engineers.

Dresel, K. M., Gardner, D. J., and Knapp, R. K. (1985). Speech compression as a potential and to automated voice warnings in the cockpit. In the Proceedings of the IEEE 1985 National Aerospace and Electronics Conference, 975-978. Dayton, Ohio: Institute of Electrical and Electronic Engineers.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (1982). Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. The Journal of the Acoustical Society of America, 71, 1568-1580.

Eberts, R. E. (1987). Internal models, tracking strategies, and dual task performance. Human Factors, 29, 407-419.

Eddy, D. R. (1987). Generic Workstation Speech Analysis and Integration, Milestone 4: Final Report. Dayton, Ohio: NTI, Inc. (Unpublished).

Eggemeier, F. T. (1984). Workload metrics for system evaluation. In the Proceedings of the Defense Research Group Panel VIII Workshop: "Application of System Ergonomics to Weapon System Development," C/5-C/20, Shriveham, England.

Eisler, F. G. (1968). Psycholinguistics: Experiments in Spontaneous Speech. New York: Academic Press.

Ellison, M. G. and Roberts, B. B. (1985). Timebased analysis of significant coordinated operations (TASCO): A cockpit workload analysis technique. In the Proceedings of Human Factors Society 29th Annual Meeting, 774-778. Santa Monica, California: Human Factors Society.

Ephrath, A. R. (1976). A novel approach to the cross adaptive auxiliary task. In the Proceedings of the Twelfth Annual Conference on Manual Control, NASA TMX-73, 63-71. Urbana-Champaign, Illinois: University of Illinois, NTIS No. N77-14746.

Eschenbrenner, A. J. (1971). Effects of intermittent noise on the performance of a complex psychomotor task. Human Factors, 13, 59-63.

Farrow, J. E. and Skerjanec, R. E. (1985). Transmission monitoring and control of strategic communication systems. In proceedings from MILCOM '85, IEEE Military Communications Conference, 485-489. Boston, Massachusetts: Institute of Electrical and Electronic Engineers.

Finkelman, J. M. and Glass, D. C. (1970). Reappraisal of the relationship between noise and human performance by means of a subsidiary task measure. Journal of Applied Psychology, 54, 211-213.

Fisk, A. D., Derrick, W. L., and Schneider, W. (1983). The asessment of workload: Dual task methodology. In the Proceedings of the Human Factors Society 27th Annual Meeting, 229-233. Santa Monica, California: Human Factors Society.

Flach, J. M., Riccio, G. E., McMillan, G. R., and Warren, R. (1986). Psychophysical methods for equating performance between alternative motion simulators. Ergonomics, 29, 1423-1438.

Fleming, R. A. (1970). The processing of conflicting information in a simulated tactical decision-making task. Human Factors, 12, 375-385.

Ford, W. R., Chapanis, A., and Weeks, G. D. (1979). Self-limited and unlimited word usage during problem solving in two telecommunication modes. Journal of Psycholinguistic Research, 8, 451-475.

Forshaw, S. E. (1981). Sidetone-level considerations in aircraft communication systems. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CCP-311, 17.1-17.5. London: Technical Editing and Reproduction Ltd.

French, N. R. and Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. The Journal of the Acoustical Society of Amercia, 19, 90-119.

Frick, F. C. and Sumby, W. H. (1952). Control tower language. The Journal of the Acoustical Society of America, 24, 595-596.

Friedman, A. and Poison, M. C. (1981). Hemispheres as independent resource systems: Limited-capacity processing and cerebral specialization. Journal of Experimental Psychology: Human Perception and Performance, 7, 1031-1058.

Friedman, A., Poison, M. C., Dafoe, C. G., and Gaskill, S. J. (1982). Dividing attention within and between hemispheres: Testing a multiple resources approach to limited-capacity information processing. Journal of Experimental Psychology: Human Perception and Performance, 8, 625-650.

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance, 6, 110-125.

Gardner, D. J., Barrett, B., Bonneau, J. R., Doucet, K., Vander Weyden, P., and Dresel, K. M. (1986). Recognition of synthesized, compressed speech in noisy environments. In the Proceedings of the Human Factors Society 30th Annual Meeting, 927-930. Santa Monica, California: Human Factors Society.

Gartner, W. B. and Murphy, M. R. (1976, November). Pilot Workload and Fatigue: A Critical Survey of Concepts and Assessment Techniques, Technical Note No. NASA TN-D-8365. Washington, D.C.: National Aeronautics and Space Administration.

Gawron, V. J. (1982). Performance effects of noise intensity, psychological set, and task type and complexity. Human Factors, 24, 225-243.

Gawron, V. J. (1984). Noise: Effect and aftereffect. Ergonomics, 27, 5-18.

Gawron, V. J. (1985). The effects of noise on decision making. In the Proceedings of the Human Factors Society 29th Annual Meeting, 1098-1101. Santa Monica, California: Human Factors Society.

Gloudemans, M. P. C. (1981). Hearing standards for aircrew. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 12.1-12.4. London: Technical Editing and Reproduction Ltd.

Goetll, B. P. and Wickens, C. D. (1984). The effect of strategy in second order manual control on resource competition with a Sternberg memory search task, Technical Report No. EPL-84-3/ONR-84-2. Arlington, Virginia: Office of Naval Research, DTIC No. ADA150985.

Gopher, D., Brickner, M., and Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. Journal of Experimental Psychology: Human Perception and Performance, 8, 146-157.

Gopher, D. and Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance (Chapter 41). New York: Wiley.

Gopher, D. and North, R. A. (1974). The Measurement of Operator Capacity by Manipulation of Dual-Task Demands, Report No. ARL-74-21/AFOSR-74-15. Savoy, Illinois: University of Illinois at Urbana-Champaign, Aviation Research Laboratory.

Grantham, D. W. and Yost, W. A. (1982). Measures of intensity discrimination. Journal of the Acoustical Society of America, 72, 406-410.

Green, R. and Flux, R. (1976). Auditory communication and workload. In the Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, A4/1-A4/8. London: Technical Editing and Reproduction Ltd.

Greene, J. H., Arbak, C. J., Courtright, J. F., and O'Donnell, R. D. (1981). Ground Attack Tactics Survey (GATS), I. Methodology Development, Technical Report No. AFAMRL-TR-81-68. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. Perception and Psychophysics, 28, 267-283.

Hamilton, J. W., Bloom,. R. F., and Bishop, E. W. (1980). Nap-of-the-earth flight: Intra-cockpit communications terms and procedures. Alexandria, Virginia: U.S. Army Research Insititute, DTIC No. ADA103157.

Harns, S. D., Owens, J. M., and North, R. A. (1978). Human performance in time-shared verbal and tracking tasks: Speech as a parallel channel for processing information and controlling systems, Report No. ZF5124004-9017. Pensacola, Florida: Naval Medical Research and Development Command.

Harris, S. D., North, R. A., and Owens, J. M. (1978). A system for the assessment of human performance in concurrent verbal and manual control tasks. Behavior Research Methods and Instrumentation, 10, 329-333.

Hart, S. G. (1975). Time estimation as a secondary task to measure workload. In the Proceedings of the Eleventh Annual Conference on Manual Control, Report No. NASA TMX-62, 64-77. Moffett Field, California: NASA Ames Research Center.

Hart, S. G. and Bortolussi, M. R. (1984). Pilot errors as a source of workload. Human Factors, 26, 545-556.

Harvey, D. S. (1988). Talking with airplanes. Air Force Magazine, 71, 88-96.

Hawkins, H. L. and Ketchum, R. D. (1980, January). The Case Against Secondary Task Analyses of Workload, Technical Report No. 6. Eugene, Oregon: University of Oregon, Center for Cognitive and Perception, DTIC No. ADA680792.

Hemingway, J. C. (1984). An Experimental Evaluation of the Sternberg Task as a Workload Metric for Helicopter Flight Handling Qualities (FHQ) Research, Technical Memorandum No. NASA-TM-85884. Moffett Field, California: Ames Research Center, NTIS No. N84-20176.

Herman, G., Ordun, M., and Riley, C. (1986). Between laboratory and field trial: Experience with a communications services testbed. In the Proceedings of the Human Factors Society 30th Annual Meeting, 804-808. Santa Monica, California: Human Factors Society.

Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. Quarterly Journal of Experimental Psychology, 22, 28-36.

Hosman, R. J. A. W. (1975). Pilot's Tracking Behavior Under Additional Workload, Report No. VTH-199. Delft, The Netherlands: Delft University of Technology, Department of Aeronautical Engineering, NTIS No. N77-13657.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. The Journal of the Acoustical Society of America, 29, 296-305.

Huchingson, R. D. (1981). Design for speech communications and training. In R. D. Huchingson, New Horizons for Human Factors in Design (Chapter 12). New York: McGraw-Hill.

Huddleston, H. F. (1974). Personality and apparent operator capacity. Perceptual and Motor Skills, 38, 1189-1190.

Jahns, D. W. (1972). Operator workload: What is it and how should it be measured? In the Proceedings of a Interagency Conference on Management and Technology in the Crew System Design Process. Los Angeles, California.

Jennings, A. E. and Chiles, W. D. (1977). An investigation of time-sharing ability as a factor in complex performance. Human Factors, 19, 535-547.

Jerison, H. J. (1959). Effects of noise on human performance. Journal of Applied Psychology, 43 , 96-101.

Jex, H. R., Jewell, W. F., and Allen, R. W. (1972, May). Development of the dual-axis and cross-coupled critical task. In the Proceedings of the Eighth Annual Conference of Manual Control, 529-552. Ann Arbor, Michigan: University of Michigan, NTIS No. N73-20028.

Johnson, D. F. and Haygood, R. C. (1984). The use of secondary tasks in adaptive training. Human Factors, 26, 105-108.

Johnston, M. E. (1975). The effect of a tracking task on speech intelligibility in noise, Technical Report No. RAE-TR-75014. Farnborough, Hants, UK: Royal Aircraft Establishment, DTIC No. ADA014516.

Jusczyk, P. W., Smith, L. B., and Murphy, C. (1981). The perceptual classification of speech. Perception and Psychophysics, 30, 10-23.

Kanagasabay, S. (1981). Hearing conservation. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 24.1-24.6. London: Technical Editing and Reproduction Ltd.

Kantowitz, B. H., Hart, S. G., and Bortolussi, M. R. (1983). Measuring pilot workload in a moving-base simulator: I. Asynchronous secondary choice-reaction task. In the Proceedings of the Human Factors Society 27th Annual Meeting. Santa Monica, California: Human Factors Society.

Kantowitz, B. H. and Weldon, M. (1985). On scaling performance operating characteristics: Caveat emptor. Human Factors, 27, 531-547.

Kelly, C. R. and Wargo, M. J. (1967). Cross-adaptive operator loading tasks. Human Factors, 9, 395-404.

Klapp, S. T., Kelly, P. A., and Netick, A. (1987). Hesitations in continuous tracking induced by a concurrent discrete task. Human Factors, 29, 327-337.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 6, 279-312.

Klemmer, E. T. (1962). Communication and human performance. Human Factors, 4, 75-79.

Kobus, D. A., Russotti, J., Schlichting, C., Haskell, G., Carpenter, S., and Wojtowicz, J. (1986). Multimodal detection and recognition performance of sonar operator. Human Factors, 28, 23-29.

Koelega, H. S. and Brinkman, J. A. (1986). Noise and vigilance: An evaluation review. Human Factors, 28, 465-482.

Koelega, H. S., Brinkman, J. A., and Bergman, H. (1986). No effect of noise on vigilance performance? Human Factors, 28, 581-594.

Kuhl, P. K. and Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Perception and Psychophysics, 32, 542-550.

Kuhl, P. K. and Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place features in macaques. The Journal of the Acoustical Society of America, 73, 1003-1010.

Lane, D. (1977). Attention allocation and the relationship between primary and secondary task difficulty. Acta Psychologica, 41, 493-495.

Lahat, M., Niederjohn, R. J., and Krubsack, D. A. (June, 1987). A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-35, 741-750.

Lazarus, H. (1986). Prediction of verbal communication in noise - A review: Part 1. Applied Acoustics, 19, 439-464.

Lazarus, H. (1987). Prediction of verbal communication in noise - A development of generalized SIL curves and the quality of communication (Part 2). Applied Acoustics, 20, 245-261.

Lea, W. A. (1980a). Speech recognition: Past, present, and future. In W. A. Lea (Ed.), Trends in Speech Recognition, 39-98. Englewood Cliffs, New Jersey: Prentice-Hall

Lea, W. A. (Ed.), (1980b). Trends in Speech Recognition. Englewood Cliffs, New Jersey: Prentice-Hall.

Levison, W. H. (1979). A model for mental workload in tasks requiring continuous information processing. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement, 189-218. New York: Plenum Press.

Loeb, M. (1986). Noise and Human Efficiency. New York: Wiley.

Luce, P. A. and Pisoni, D. B. (1983). Capacity-Demanding Encoding of Synthetic Speech in Serial-Ordered Recall. Research on Speech Perception, Progress Report No. 9, 295-309. Bloomington, Indiana: Indiana University.

Lynch, J. T. (1987). A methodology for evaluating the performance of dynamic range control algorithms for speech enhancement. In the Proceedings: ICASSP 87. 1987 International Conference on Acoustics, Speech, and Signal Processing, 153-156. Dallas, Texas: Institute of Electrical and Electronic Engineers.

McCauley, M. E. (1984). Human factors in voice technology. In F. A. Muckler (Ed.), Human Factors Review, 131-166. Santa Monica, California: Human Factors Society.

McCollom, I. N. and Chapanis, A. (1956). A Human Engineering Bibliography, Technical Report No. 15. San Diego, California: Arts and Crafts Press.

McKinley, R. L., Anderson, T. A., and Moore, T. J. (1982). Evaluation of speech synthesis for use in military noise environments. In the Proceedings of the Workshop on Standardization for Speech I/O Technology, 241-245. Gaithersburg, Maryland: National Bureau of Standards.

McKinley, R. L. and Kurtz, R. V. (1984). Subjective performance of selected speech codes in the presence of channel errors. In the Proceedings of the IEEE Military Communications Conference: MILCOM '84 (Vol. 1). Los Angeles, California: Institute of Electrical and Electronic Engineers.

McKinley, R. L., Nixon, C. W., and Moore, T. J. (1981). Voice communications capability of selected in-flight headgear devices. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 19.1-19.8. London: Technical Editing and Reproduction Ltd.

Makhoul, J. (1985). Automatic speech recognition in severe environments: A report on the National Research Council study. In the Proceedings of Speech Tech '85, 228-229. New York: Media Dimensions.

Marchetti, F. M. and Tsuji, B. H. (1986). Integrated voice-and-data user interface designs. In the Proceedings of the Human Factors Society 30th Annual Meeting, 814-818. Santa Monica, California: Human Factors Society.

Mayer, M. S. and Lindburg, A. W. (1981). Electronic voice communications improvements for Army aircraft. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 5.1-5.12. London: Technical Editing and Reproduction Ltd.

Metzler, T. R. and Shingledecker, C. A. (1982). Register of Research in Progress on Mental Workload, Technical Report No. AFAMRL-TR-82-42. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Micallizzi, J. (1981). The Structure of Processing Resource Demands in Monitoring Automatic Systems. Unpublished master's thesis, University of Illinois, Urbana-Champaign, Illinois.

Micallizzi, J. and Wickens, C. D. (1980). The Application of Additive Factors Methodology Workload Assessment in a Dynamic SysteMonitoring Task, Interim Technical Report No. EPL-80-2/ONR-80-2. Arlington, Virginia: Office of Naval Research, DTIC No. ADA094314.

Michaelis, P. R. (1984). Fundamentals of speech technology. In the Proceedings of the Human Factors Society 28th Annual Meeting, 552-554. Santa Monica, California: Human Factors Society.

Michon, J. A. (1966). Tapping regularity as a measure of perceptual motor load. Ergonomics, 9, 401-412.

Miller, G. A., Heise, G. A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. Journal of Experimental Psychology, 41, 329-335.

Mills, C. B. (1980). Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. Journal of Experimental Psychology: Human Perception and Performance, 6, 528-535.

Mirabella, A. and Goldstein, D. A. (1967). The effects of ambient noise upon signal detection. Human Factors, 9, 277-284.

Mirchandani, P. B. (1972). An auditory display in a dual-axis tracking task. IEEE Transactions on Systems, Man, and Cybernetics, SMC-2, 375-380.

Møller, A. R. (1962). Acoustic reflex in man. The Journal of the Acoustical Society of America, 34, 1524-1534.

Money, K. E. (Ed.). (1981). Aural Communication in Aviation, AGARD-CP-311. Twenty-seven papers presented at the Aerospace Medical Panel Specialists' Meeting held at Soesterberg Air Base, Netherlands, from 30 March to 2 April 81. London: Technical Editing and Reproduction Ltd. (North Atlantic Treaty Organization [NATO], Advisory Group for Aerospace Research and Development [AGARD], Neuilly Sur Seine, France).

Moore, T. J. (1979). Recommended Test Procedures for Performing Relative Psychoacoustical Evaluations of Seek Talk Radio Systems, Technical Report No. AMRL-TR-79-40. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory.

Moore, T. J., Nixon, C. W., and McKinley, R. L. (1981). Comparative intelligibility of speech materials processed by standard Air Force communication systems in the presence of simulated cockpit noise. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 14.1-14.5. London: Technical Editing and Reproduction Ltd.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance (Chapter 40). New York: Wiley.

Mosko, J. D. (1981). "Clear Speech": A strategem for improving radio communications and automatic speech recognition in noise. In K. E. Money (Ed.), Aural Communication in Aviation, GARD-CP-311, 4.1-4.6. London: Technical Editing and Reproduction Ltd.

Mowbray, G. H. (1964). Peception and retention of verbal information presented during auditory shadowing. The Journal of the Acoustical Society of America, 36, 1459-1464.

Nixon, C. E. (1979). Hearing protective devices: Ear protectors. In C. M. Harris (Ed.), Handbook of Noise Control (Chapter 12). New York: McGraw-Hill.

Nixon, C. W. and Decker, W. H. (1985). Voice Communications Effectiveness of the All-Purpose MCU-2/P Chemical Defense Protective Mask, Technical Report No. AAMRL-TR-85-050. Wright-Patterson Air Force Base, Ohio: Armstrong Aerospace Medical Research Laboratory, DTIC No. ADA161031.

Nygren, T. E. (1982). Conjoint Measurement and Conjoint Scaling: A Users Guide, Technical Report No. AFAMRL-TR-82-22. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

O'Donnell, R. D. (1982). Historical foundations of the AFAMRL workload program. In the Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Technical Report No. AFFTC-TR-82-5. Edwards Air Force Base, California: Air Force Flight Test Center.

Ogden, G. D., Anderson, N. S., and Reick, A. M. (1979). Dual task measures of S-R compatibility. In the Proceedings of the Human Factors Society 23rd Annual Meeting. Santa Monica, California: Human Factors Society.

Orrell, D. H., II. (1985). Effects of Noise and Workload on a Communication Task (Thesis Proposal), Technical Report No. AFIT/CI/NR 85-101T. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, DTIC No. ADA160743.

Owens, J. M., Goodman, L. S., and Pianka, M. J. (1984). Interference Effects of Vocalization on Dual Task Performance, Report No. NAMRL-1309. Pensacola, Florida: Naval Aerospace Medical Research Laboratory, DTIC No. ADA150687.

Parasuraman, R. (1981). Vigilance, monitoring, and speech. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook Perception and Human Performance, Vol. II, Cognitive Processes and Performance (Chapter 43). New York: Wiley.

Parks, D. L. (1979). Current workload methods and emerging challenges. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement, 387-416. New York: Plenum Press.

Pastore, R. E. and Sorkin, R. D. (1972). *Simultaneous two-channel signal detection.* I. Simple binaural stimuli. The Journal of the Acoustical Society of America, 51, 544-551.

Pavlovic, C. V. (1987). Derivation of primary parameters and procedures for use in speech intelligibility predictions. The Journal of the Acoustical Society of America, 82, 413-422.

Pearson, W. H. and Shew, R. L. (1980). Studies in Tactical Symbology, II. Symbol Meaningfulness and Learning Efficiency, Technical Report No. AFAMRL-TR-80-115. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Pew, R. W. (1979). Secondary tasks and workload measurement. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement, 23-28. New York: Plenum Press.

Pickett, J. M. (1980). The Sounds of Speech Communication. A Primer of Acoustic Phonetics and Speech Perception. Baltimore, Maryland: University Park Press.

Pisoni, D. B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. The Journal of the Acoustical Society of America, 70, RR10 (Abstract from 102nd Meeting of ASA).

Pisoni, D. B. (1986). A brief overview of speech synthesis and recognition technologies. In the Proceedings of the Human Factors Society 30th Annual Meeting, 1326-1330. Santa Monica, California: Human Factors Society.

Pisoni, D. B. and Koen, E. (1982). Intelligibility of natural and synthetic speech at several different signal-to-noise ratios. The Journal of the Acoustical Society of America, 71, VV I (Program Abstract of 103th Meeting of ASA).

Poock, G. K. (1986). Speech recognition: Research, applications and international efforts. In the Proceedings of the Human Factors Society 30th Annual Meeting, 1278-1283. Santa Monica, California: Human Factors Society.

Power, M. J. (1986). A technique for measuring processing load during speech production. Journal of Psycholinguistic Reseach, 15, 371-381.

Pratt, R. L. (1985). Assessing the intelligibility and acceptability of voice communciation systems. In the Proceedings of MILCOM '85, IEEE Military Communications Conference, 176-180. Boston, Massachusetts: Institute of Electrical and Electronic Engineers.

Rahimi, M. and Wierwille, W. W. (1982). Evaluation of the sensitivity and intrusion of workload estimation techniques in piloting tasks emphasizing mediational activity. In the Proceedings of the 1982 IEEE International Conference on Cybernetics and Society, 593-597. New York: Institute of Electrical and Electronic Engineers.

Rehmann, J. T., Stein, E. S., and Rosenberg, B. L. (1983). Subjective pilot workload assessment. Human Factors, 25, 297-302.

Robinson, C. P. and Eberts, R. E. (1987). Comparison of speech and pictorial displays in a cockpit environment. Human Factors, 29, 31-44.

Rolfe, J. M., Cook, J. R., and Durose, C. G. (1986). Knowing what we get from training devices: Substituting a little arithmetic for a little emotion. Ergonomics, 29, 1415-1422.

Rollins, H. A., Jr. and Hendricks, R. (1980). Processing of words presented simultaneously to eye and ear. Journal of Experimental Psychology: Human Perception and Performance, 6, 99-109.

Rubenstein, H., Garfield, L., and Millikan, J. A. (1970). Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 9, 487-494.

Sanders, A. F. (1979). Some remarks on mental load. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press.

Savage, R. E., Wierwille, W. W., and Cordes, R. E. (1978). Evaluating the sensitivity of various measures of operator workload using random digits as a secondary task. Human Factors, 20, 649-654.

Scharf, B. and Buus, S. (1986). Audition I: Stimulus, physiology, thresholds. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. I, Sensory Processes and Perception (Chapter 14). New York: Wiley.

Scharf, B. and Houtsman, A. J. M. (1986). Audition II: Loudness, pitch, localization, aural distortion, pathology. In K. P. Boff, L. Kaufman, and J. P. Thomas (Eds.), Human of Perception and Human Performance, Vol I, Sensory Processes and Perception (Chapter 15). New York: Wiley.

Schiflett, S. G. (1980). Evaluation of a Pilot Workload Assessment Device to Test Alternate Display Formats and Control Handling Qualitities, Final Report No. SY-33R-80. Patuxent River, Maryland: Naval Air Test Center.

Schlegel, R. E., Gilliland, K., and Schlegel, B. (1986). Development of the criterion task set performance data base. In the Proceedings of the Human Factors Society 30th Annual Meeting, 58-62. Santa Monica, California: Human Factors Society.

Schori, T. R. (1973). A comparison of visual, auditory, and cutaneous tracking displays when divided attention is required to a cross-adaptive loading task. Ergonomics, 16, 153-158.

Schwab, E. C., Nusbaum, H. C., and Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. Human Factors, 27, 395-408.

Shaffer, M. T., Shafer, J. B., and Kutche, G. B. (1986). Emperical workload and communications: Analysis of Scout helicopter exercises. In the Proceedings of the Human Factors Society 30th Annual Meeting, 628-632. Santa Monica, California: Human Factors Society.

Shannon, C. E. (1948). The mathematical theory of communication. Bell System Technical Journal, 301-344.

Sheridan, T. B. (1980). Mental workload - What is it? Why bother with it? Human Factors Society Bulletin, 23, 1-2.

Shingledecker, C. A. (1980a). Enhancing operator acceptance and noninterference in secondary task measures of workload. In the Proceedings of the Human Factors Society 24th Annual Meeting, 674-677. Santa Monica, California: Human Factors Society.

Shingledecker, C. A. (1980b). Operator Strategy: A Neglected Variable in Workload Assessment. Paper presented at the 88th Annual Convention of the American Psychological Association. Montreal, Quebec, Canada.

Shingledecker, C. A., Acton, W. H., and Crabtree, M. S. (1983). Development and Application of a Criterion Task Set for Workload Metric Evaluation. Paper presented at the Second Aerospace Behavioral Engineering Technology Conference, Long Beach, California.

Shingledecker, C. A. and Crabtree, M. S. (1980). Analytical Scaling of the Workload of Secondary Communications Tasks, Technical Report No. SRL-TSG-TM-80-02. Dayton, Ohio: Systems Research Laboratories, Inc. (Unpublished).

Shingledecker, C. A., Crabtree, M. S., Simons, J. C., Courtright, J. F., and O'Donnell, R. D. (1980). Subsidiary Radio Communications Task for Workload Assessment in RandD Simulations: I. Task Development and Workload Scaling, Technical Report No. AFAMRL-TR-80-126. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. W. (1986). Group processes in computer-mediated communication. Organizational Behavior and Human Decision Process, 37, 157-187.

Simons, J. (1967). Low-Altitude RECCE/Strike Techniques, Problems, Technical Report No. ASD-TR-67-17. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division.

Simons, J. C. and Craig, J. L. (1983). Electroluminescent Formation Lights for HC-130 P/N Special Operations: I. Flight Test Candidates , Technical Report No. AFAMRL-TR-83-069. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Simons, J. C. and Valverde, H. H. (1971). Voice Communications Training for Forward Air Controller/Strike Target Location, Technical Research Memorandom No. AFHRL-(TR)-TRM-2. Wright-Patterson Air Force Base, Ohio: Air Force Human Resources Laboratory.

Simpson, C. A. (1985). Selecting cockpit functions for speech I/O technology. In the Proceedings of the IEEE 1985 National Aerospace and Electronics Conference, 979-986. Dayton, Ohio: Institute of Electrical and Electronic Engineers.

Simpson, C. A. (1986). Speech variability effects on recognition accuracy associated with concurrent task performance by pilots. Ergonomics, 29, 1343-1357.

Simpson, C. A., McCauley, M. E., Roland, E. F., Ruth, J. C., and Williges, B. H. (1985). System design for speech recognition and generation. Human Factors, 27, 115-142.

Simpson, C. A. and Marchionda-Frost, K. (1984). Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots. Human Factors, 26, 509-517.

Simpson, C. A., Moore, C. A., and Ruth, J. C. (1986). Testing for robust speech recognition performance. In the Proceedings of the IEEE 1986 National Aerospace and Electronics Conference (NAECON 1986), 790-798. Dayton, Ohio: Institute of Electrical and Electronic Engineers.

Sims, J. T. (1985). A speech-to-noise ratio measurement of algorithm. The Journal of the Acoustical Society of America, 78, 1671-1674.

Singleton, W. T., Fox, J. G., and Whitfield, D. (1971). Measurement of Man at Work: An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. London: Taylor and Francis Ltd.

Slaby, C. A., Hartley, C. S., and Pulliam, R. (1985). Speech recognition, speech synthesis and touch-panel technology applied to the control of remotely piloted spacecraft. In the Proceedings of the Human Factors Society 29th Annual Meeting, 1140-1143. Santa Monica, California: Human Factors Society.

Smith, H. P. R. (1979). A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions , Technical Memorandum No. NASA-TM-78482. Moffett Field, California: NASA Ames Research Center.

Smoorenburg, G. F., de Laat, J. A. P. M., and Plomp, R. (1981). The effect of noise-induced hearing loss on the intelligibility of speech in noise. In K. E. Money (Ed.), Aural Communicatoin in Aviation, AGARD-CP-311, 11.1-11.8. London: Technical Editing and Reproduction Ltd.

Sorkin, R. D., Pastore, R. E., and Pohlmann, L. D. (1972). Simultaneous two-channel signal detection. II. Correlated and uncorrelated signals. The Journal of the Acoustical Society of America, 51, 1960-1965.

Sorkin, R. D., Pohlmann, L. D., and Gilliom, J. D. (1973). Simultaneous two-channel signal detection. III. 630- and 1400-Hz signals. The Journal of the Acoustical Society of America, 53, 1045-1050.

Sorkin, R. D., Pohlmann, L. D., and Woods, D. D. (1976). Decision interaction between auditory channels. Perception and Psychophysics, 19, 290-295.

Sperling, G., and Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. I, Sensory Processes and Perception (Chapter 2). New York: Wiley.

Spicuzza, R. J. and Crabtree, M. S. (1980). Automated Item Recognition as a Secondary Task for Objective Assessment of Aircrew Workload in Flight Simulators. Dayton, Ohio: Systems Research Laboratories, Inc. (Unpublished paper).

Spicuzza, R., Pinkus, A., and O'Donnell, R. D., (1974). Development of Performance Assessment Methodology for the Digital Avionics Information System. Dayton, Ohio: Systems Research Laboratories, Inc. (Unpublished report).

Steeneken, H. J. M. and Langhout, G. (1986). Voice interactive systems in severe noise conditions. In the AGARD Conference Proceedings on Information Management and Decision Making in Advacned Airborne Weapon Systems, AGARD-CP-414. Neuilly Sur Seine, France: Advisory Group for Aerospace Research and Development.

Stoll, F. C., Hoecker, D. G., Krueger, G. P., and Chapanis, A. (1976). The effects of four communication modes on the structure of language used during cooperative problem solving. The Journal of Psychology, 94,13-26.

Swets, J. A. (Ed.). (1964). Signal Detection and Recognition by Human Observers: Contemporary Readings. New York: Wiley.

Systems Research Laboratories, Inc. (1980). Design and Fabrication of a Workload Assessment Device for an Adaptive, Cross-Coupled, Airborne Version of the NATC WAD Item Recognition Task, SRL Technical Proposal No. 21269. Dayton, Ohio: Systems Research Laboratories, Inc.

Systems Research Laboratories, Inc. (1981). Wordload Assessment Methods (CBD Response). Dayton, Ohio: Systems Research Laboratories, Inc.

Systems Research Laboratories, Inc. (1986). Controlling Operator Workload in Army System Design and Evaluation, SRL Technical Proposal No. 25687. Dayton, Ohio: Systems Research Laboratories, Inc. (Unpublished).

Teichner, W. H., Arees, E., and Reilly, R. (1963). Noise and human performance. A psychophysiological approach. Ergonomics, 6, 83-97.

Thomas, J. C., Rosson, M. B., and Chodorow, M. (1984). Human factors and synthetic speech. In the Proceedings of the Human Factors Society 28th Annual Meeting. Santa Monica, California: Human Factors Society.

Thurlow, W. (1957). An auditory figure-ground effect. The American Journal of Psychology, 70, 653-654.

Thwing, E. J. (1956). Effect of repetition on articulation scores for PB words. The Journal of the Acoustical Society of America, 28, 302-303.

Tsang, P. S. (1986). Display/control integrality and time-sharing performance. In the Proceedings of the Human Factors Society 30th Annual Meeting, 445-449. Santa Monica, California: Human Factors Society.

Tsang, P. S. and Rothschild, R. A. (1985). To speak or not to speak: A multiple resource perspective. In the Proceedings of the Human Factors Society 29th Annual Meeting, 76-80. Santa Monica, California: Human Factors Society.

Tsang, P. S. and Wickens, C. D. (1984). The effects of task structures on time-sharing efficiency and resource allocation optimality. In the Proceedings of the 20th Annual Conference on Manual Control, Vol. 2, 305-318. Washington, D.C.: National Aeronautics and Space Administration.

Vidulich, M. A. (1986). Response modalities and time-sharing performance. In the Proceedings of the Human Factors Society 30th Annual Meeting, 337-341. Santa Monica, California: Human Factors Society.

Vidulich, M. A. and Tsang, P. S. (1986). Techniques of subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods. Ergonomics, 29, 1385-1398.

Vidulich, M. and Wickens, C. D. (1981). Time-Sharing Manual Control and Memory Search: The Joint Effects of Input and Output Modality Competition, Priorities and Controlled Order, Technical Report No. EPL-81-4/ONR-81-4. Arlington, Virginia: Office of Naval Research, DTIC No. ADA108721.

Vinje, E. W. and Pitkin, E. T. (1972a). Human operator dynamics for aural compensatory tracking. IEEE Transactions on Systems, Man, and Cybernetics, SMC-2, 504-512.

Vinje, E. W. and Pitkin, E. T. (1972b). Human operator dynamics for aural compensatory tracking. In the Proceedings of the Seventh Annual Conference on Manual Control, NASA-SP-281, 339-348. Washington, D.C.: National Aeronautics and Space Administration.

Warm, J. S. and Parasuraman, R. (1987). Vigilance: Basic and applied research (Special Issue). Human Factors, 29, 623-755 (whole volume).

Warren, R. M. and Warren, R. P. (1970). Auditory illusions and confusions. Scientific American, 223, 30-36.

Way, T. C. (1978). Verification of Workload - A Job for Simulation. Washington, D.C.: American Institute of Aeronautics and Astronautics.

Weeks, G. D., Kelly, M. J., and Chapanis, A. (1974). Studies in interactive communication: V. Cooperative problem solving by skilled and unskilled typists in a teletypewriter mode. Journal of Applied Psychology, 59, 665-674.

Welford, A. T. (1978). Mental workload as a function of demand, capacity, strategy and skill. Ergonomics, 21, 151-167.

Wempe, T. E. and Baty, D. L. (1968). Human information processing rates during certain multi-axis tracking tasks with a concurrent auditory task. IEEE Transactions on Man-Machine systems, MMS-9, 129-138.

Wenzel, E. M. (1986). Effects of stimulus characteristics and task demands on pilots' perception of dichotic messages. In the Proceedings of the Human Factors Society 30th Annual Meeting, 1177-1181. Santa Monica, California: Human Factors Society.

Werkowitz, E. (1980). Ergonomic considerations for the cockpit applications of speech generation technology. In the Proceedings of the Voice Interactive Systems Society: Applications and Payoffs, 295-309. Moffett Field, California: NASA Ames Research Center.

Wewerinke, P. H. (1974). Human Operator Workload for Various Control Situations. Amsterdam, The Netherlands: National Aerospace Laboratory NLR. (Also in the Proceedings of the 10th Annual Conference on Manual Control, 1974, 167-192. Wright-Patterson Air Force Base, Ohio.)

Whitaker, L. A. (1979). Dual-task interference as a function of cognitive processing load. Acta Psychologica, 43, 71-84.

Wickens, C. D. (1979). Measures of workload, stress and secondary tasks. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement, 79-99. New York: Plenum Press.

Wickens, C.D. (1988, January). Multiple Resource Theory: Its Strengths and Limitations. Colloquium presentation at the University of Dayton, Dayton, Ohio.

Wickens, C. D. and Goettl, B. P. (1984). The Effect of Strategy in Second Order Manual Control on Resource Competition with a Sternberg Memory Search Task, Technical Report No. EPL-84-3/ONR-84-2. Arlington, Virginia: Office of Naval Research.

Wickens, C. D., Hyman, F., Dellinger, J., Taylor, H., and Meador, M. (1986). The Sternberg memory search task as an index of pilot workload. Ergonomics, 29, 1371-1384.

Wickens, C. D., Mountford, S. J., and Schreiner, W. (1981). Multiple resources, task-hemispheric integrity, and individual differences in time-sharing. Human Factors, 23, 211-229.

Wickens, C. D. and Sandry, D. L. (1980). An Investigation of the Dual Task Performance Relationship to Hemispheric Specialization, Technical Report No. EPL-80-1/ONR-80-1). Arlington, Virginia: Office of Naval Research, DTIC No. ADA086808.

Wickens, C. D., Sandy, D. L., and Hightower, R. (1982). Display Location of Verbal and Spatial Material: The Joint Effects of Task-Hemispheric Integrity and Processing Strategy, Technical Report No. EPL-82-2/ONR-82-2. Arlington, Virginia: Office of Naval Research, DTIC No. ADA121000.

Wickens, C. D., Sandry, D. L., and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. Human Factors, 25, 227-248.

Wickens, C. D. and Vidulich, M. (1982). S-C-R- Compatibility and Dual Task Performance in Two Complex Information Processing Tasks: Threat Evaluation and Fault Diagnosis, Technical Report No. EPL-82-3/ONR-82-3. Champaign, Illinois: University of Illinois, Engineering Psychology Laboratory.

Wickens, C. D., Vidulich, M., and Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. Human Factors, 26, 533-543.

Wickens, C. D. and Yeh, Y. (1985). POCs and performance decrements: A reply to Kantowitz and Weldon. Human Factors, 27, 549-554.

Wickens, C. D., Zenyuh, J., Culp, V., and Marshak, W. (1985). Voice and manual control in dual task situations. In the Proceedings of the Human Factors Society 29th Annual Meeting, 1093-1097. Santa Monica, California: Human Factors Society.

Wiener, E. L., Poock, G. K., and Steele, M. (1964). Effect of time-sharing on monitoring performance: Simple mental arithmetic as a loading task. Perceptual and Motot Skills, 19, 435-440.

Wierwille, W. W. and Gutmann, J. C. (1978). Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. Human Factors, 20, 233-244.

Wierwille, W. W., Rahimi, M., and Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. Human Factors, 27, 489-502.

Wierwille, W. W. and Williges, B. H. (1980, March). An Annotated Bibliography on Operator Mental Workload Assessment, Technical Report No. SY-27R-80. Patuxent River, Maryland: Naval Air Test Center.

Wilkins, P. A. (1981). Assessing the effectiveness of auditory warnings. In K. E. Money (Ed.), Aural Communication in Aviation , AGARD-CP-311, 8.1-8.13. London: Technical Editing and Reproduction Ltd.

Williams, D. H. and Simpson, C. A. (1976). A Systematic Approach to Advanced Cockpit Warning Systems for Air Transport Operations: Line Pilot Preferences, NASA SP-416. Washington, D.C.: National Aeronautics and Space Administration.

Williams, W. E. and Young, D. D. (1967). An evaluation of four methods of monitoring simultaneous primary and secondary voice messages. Human Factors, 9, 45-52.

Zeitlin, L. R. and Finkelman, J. M. (1969). A subsidiary task evaluation of the information processing load in a vehicular control simulation. American Psychological Association Experimental Publication System, 3, Manuscript No. 106b.

Zeitlin, L. R. and Finkelman, J. M. (1975). Research note: Subsidiary task techniques of digit generation and digit recall as indirect measures of operator loading. Human Factors, 17, 218-220.

Zoltan-Ford, E. (1984). Reducing variability in natural-language interactions with computers. In the Proceedings of the Human Factors Society 28th Annual Meeting, 768-772. Santa Monica, California: Human Factors Society.

Zwicker, E. and Terhardt, E. (1979). Automatic speech recognition using psychoacoustic models. The Journal of the Acoustical Society of America, 65, 487-498.

REFERENCES

Acton, W. H., and Crabtree, M. S. (1985). User's Guide for the Criterion Task Set, Technical Report No. AAMRL-TR-85-034. Wright-Patterson Air Force Base, Ohio: Armstrong Aerospace Medical Research Laboratory.

ANSI (1960). USA Standard Method for Measurement of Monosyllabic Word Intelligibility, No. USAS-S3.2-1960. New York: American National Standards Institute, Inc.

ANSI (1969). American National Standard Methods for the Calculation of the Articulation Index, No. ANSI-S3.5-1969. New York: American National Standards Institute, Inc.

Arnoult, M. D. and Voorhees, J. W. (1980). Effects of aircraft noise on an intelligibility task. Human Factors, 22, 183-188.

Arnoult, M. D., Voorhees, J. W., and Gilfillan, L. G. (1986). Effects of simulated helicopter cabin noise on intelligibility and annoyance. Bulletin of the Psychonomic Society, 24,115-117.

Bell, P. A. (1978). *Effects of noise and heat stress on primary and subsidiary task performance.* Human Factors, 20, 749-752.

Bennighof, R. H., Farris, H. W., Lauderdale, L. K., Richard, R. H., and Wild, T. A. (1987). Effectiveness of jamming signals. In J. A. Boyd, D. B. Harris, D. D. King, and H. W. Welch, Jr. (Eds.), Electronic Countermeasures (Chapter 14). Los Altos, California: Peninsula Publications.

Beranek, L. L. (1949). Acoustic Measurements, New York: Wiley.

Boff, K. R., Kaufman, L., and Thomas, J. P. (Eds.). (1986). Handbook of Perception and Performance. Vol. II. Cognitive Processes and Performance. New York: Wiley.

Bratfisch, O. (1972). Experienced Intellectual Activity and Perceived Difficulty of Intelligence Tests, Report No. 30. Stockholm, Sweden: University of Stockholm, Institute of Applied Psychology.

Bratfisch, O., Borg, G., and Dornic, S. (1972). Perceived Item Difficulty in Three Tests of Intellectual Performance Capacity, Report No. 29. Stockholm, Sweden: University of Stockholm, Institute of Applied Psychology.

Brecht, M. (1977, August). Cardiac Arrhythmia and Secondary Tasks as Measure of Mental Load. Unpublished master's thesis, California State University at Northridge, Northridge, California.

Broadbent, D. E. (1958). Perception and Communication. New York: Pergamon Press.

Broadbent, D. E. and Gregory, M. (1964). Stimulus set and response set: The alternation of attention. Quarterly Journal of Experimental Psychology, 16, 309-318.

Brown, I. D. (1964). The measurement of perceptual load and reserve capacity. Transactions of the Association of Industrial Medical Officers, 14, 44-49.

Burke, M. W., Gilson, R. D., and Jagacinski, R. J. (1980). Multimodal information processing for visual workload relief. Ergonomics, 23, 961-975.

Callaway, E., Tueting, P., and Koslow, S. (Eds.). (1978). Brain Event-Related Potentials in Man. New York: Academic Press.

Carterette, E. C. and Friedman, M. P. (Eds.). (1976). Handbook of Perception. Vol. VIII, Language and Speech. New York: Academic Press.

Casali, J. G. and Wierwille, W. W. (1982). A sensitivity/intrusion comparison of mental workload estimation techniques using a flight task emphasizing perceptual piloting activities. In the Proceedings of the 1982 International Conference on Cybernetics and Society, 598-602. New York: Institute of Electrical and Electronic Engineers.

Casali, J. G. and Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 25, 623-641.

Chapanis, A. (1959). Research Techniques in Human Engineering (Chapter 8, Articulation Test Methods). Baltimore, Maryland: John Hopkins Press.

Chapanis, A. (1965). Man-Machine Engineering (Chapter 4, Speech Communication Systems). Belmont, California: Wadsworth.

Chapanis, A. (1971). Prelude to 2001: Explorations in human communication. American Psychologist, 21, 949-961.

Chapanis, A. (1975). Interactive human communications. Scientific American, 232, 36-42.

Chapanis, A. (1976). Interactive Human Communications: Some Lessons Learned from Laboratory Experiments, Technical Report No. 5, John Hopkins University, ONR Contract No. N00014-75-C-0131. Arlington, Virginia: Office of Naval Research.

Chapanis, A., Ochsman, R. B., Parrish, R. N., and Weeks, G. D. (1972). Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. Human Factors, 14, 487-509.

Chapanis, A. and Overbey, C. M. (1974). Studies in interactive communication: III. Effects of similar and dissimilar communication channels and two interchange options on team problem solving. Perceptual and Motor Skills, 38, 343-374.

Chapanis, A., Parrish, R. N., Ochsman, R. B., and Weeks, G. D. (1977). Studies in interactive communication: II. The effects of four communication modes on the linguistic performance of teams during cooperative problem solving. Human Factors, 19, 101-126.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. The Journal of the Acoustical Society of America, 25, 975-979.

Cherry, E. C. (1966). On Human Communication. Cambridge, Massachusetts: M.I.T. Press.

Chiles, W. D. (1982). Workload, task, and situational factors as modifiers of complex human performance. In E. A. Allusi and E. A. Fleishman (Eds.), Human Performance and Productivity. Hillsdale, New Jersey: Erlbaum Press.

Cole, R. A. (1980). Perception and Production of Fluent Speech. Hillsdale, New Jersey: Lawrence Erlbaum.

Cooper, G. E. and Harper, R. P., Jr. (1969). The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, Report No. NASA TN-D-5153. Moffitt Field, California: NASA Ames Research Center.

Deutsch, J. A. and Deutsch, D. (1963). Attention: Some theoretical considerations. Psychological Review, 70, 80-90.

Dorfman, P. W. and Goldstein, I. L. (1975). The effects of task coherency, preview, and speed-stress in timing and anticipation. Journal of Motor Behavior, 7, 45-55.

Dornic, S. (1980). Language dominance, spare capacity, and perceived effort in bilinguals. Ergonomics, 23, 366-377.

Dougherty, D. J., Emery, J. H., and Curtin, J. G. (1964). Comparison of Perceptual Workload in Flying Standard Instrumentation and the Contact Analog Vertical Display, Report No. JANAIR D 228-421-019. Fort Worth, Texas: Bell Helicopter Company.

Doyle, J. C., Ornstein, R., and Galin, D. (1974). Lateral specialization of cognitive mode: II. EEG frequency analysis. Psychophysiology, 11, 567-578.

Egan, J. P. (1948). Articulation testing methods. Laryngoscope, 58, 955-991.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (1954). Some factors affecting multichannel listening. The Journal of the Acoustical Society of America, 26, 774-782.

Eggemeier, F. T. (1981). Development of a secondary task workload assessment battery. In the Proceedings of the IEEE International Conference on Cybernetics and Society, 410-414. New York: Institute of Electrical and Electronic Engineers.

Fairbanks, G. (1958). Test of phonomeic differentiation: The rhyme test. The Journal of the Acoustical Society of America, 30, 596-600.

Finkelman, J. M., Zeitlin, L. R., Filippi, J. A., and Friend, M. A. (1977). Noise and driver performance. Journal of Applied Psychology, 62, 713-718.

Fitts, P. M. and Posner, M. I. (1967). Human Performance. Belmont, California: Brooks/Cole.

Fröhlich, G. R. (1981). The effects of ear protectors and hearing losses on sentence intelligibility in aircraft noise. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 16.1-16.2. London: Technical Editing and Reproduction Ltd.

Gevins, A. (1983). Brain potential evidence for lateralization of higher cognitive functions. In J. B. Healige (Ed.), Asymmetry: Method, Theory, Application. New York: Praeger.

Goldstein, I. L. and Dorfman, P. W. (1978). Speed and load stress as determinants in a time-sharing task. Human Factors, 20, 603-609.

Gomer, F. E., Spicuzza, R. J., and O'Donnell, R. D. (1976). Evoked-potential correlates of visual item recognition during memory-scanning tasks. Physiological Psychology, 4, 61-65.

Greene, B. G., Logan, J. S., and Pisoni, D. B. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. Behavior Research Methods, Instruments, and Computers, 18, 100-107.

Hallsten, L. and Borg, G. (1975). Six Rating Scales for Perceived Difficulty, Report No. 58. Stockholm, Sweden: University of Sweden, Institute of Applied Psychology.

Harris, C. M. (1979). Handbook of Noise Control (2nd ed.). New York: McGraw-Hill.

Hart, S. G. (1982). Theoretical basis for workload assessment research at NASA Ames Research Center. In the Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Report No. AFFTC-TR-82-5. Edwards Air Force Base, California: Air Force Flight Test Center.

Hassett, J. (1978). A Primer of Psychophysiology. San Francisco, California: Freeman.

Hawkins, H. L. and Presson, J. C. (1977). Masking level and attention in auditory recognition. In S. Dornic (Ed.), Attention and Performance VII. Hillsdale, New Jersey: Laurence Erlbaum.

Hawkins, H. L. and Presson, J. C. (1986). Auditory information processing. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. II. Cognitive Processes and Performance (Chapter 26). New York: Wiley.

Hess, R. A. (1977). Prediction of pilot opinion ratings using an optimal pilot model. Human Factors, 19, 459-476.

Hicks, T. G. and Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. Human Factors, 21, 129-143.

Hougast, T. and Steeneken, H. J. M. (1981). STI - An objective measure for the performance of voice communication systems. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311. 17.1-17.6. London: Technical Editing and Reproduction Ltd.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. The Journal of the Acoustical Society of America, 37, 158-166.

Isreal, J. B., Chesney, G. L., Wickens, C. D., and Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 17, 259-273.

Jusczyk, P. W. (1986). Speech perception. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Performance, Vol. II, Cognitive Processes and Performance (Chapter 27). New York: Wiley.

Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, New Jersey: Prentice-Hall.

Kalikow, D. N. and Stevens, K. N. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. The Journal of the Acoustic Society of America, 61, 1337-1351.

Knowles, W. B. (1963). Operator loading tasks. Human Factors, 5, 155-161.

Kraus, E. F. and Roscoe, S. N. (1972). Reorganization of airplane manual flight control dynamics. In the Proceedings of the 16th Annual Meeting of the Human Factors Society, 117-126. Santa Monica: Human Factors Society.

Kryter, K. D. (1972). Speech communication. In H. P. Van Cott and R. G. Kinkade (Eds.), Human Engineering Guide to Equipment Design (rev. ed., Chapter 5). Washington, D.C.: U.S. Government Printing Office.

Kryter, K. D. (1984). Physiological, Psychological, and Social Effects of Noise, 57-111, NASA Ref. Pub. No. 1115. Washington, D.C.: National Aeronautics and Space Administration.

Kryter, K. D. (1985). The Effects of Noise on Man (2nd ed.). New York: Academic Press.

Kryter, K. D. and Williams, C. E. (1965). Masking of speech by aircraft noise. The Journal of the Acoustical Society of America, 39, 38-150.

Lacey, L. A. (1973). Voice Advisory Intelligibility in a Simulated Dogfight Acoustic Environment, Technical Note No. 4011-16. China Lake, California: Naval Weapons Center.

Lawrence, G. H. (1979). Brain wages and the enhancement of pilot performance. In B. O. Hartman and R. E. McKenzie (Eds.), Survey of Methods to Assess Workload, AGARD-AG-246, 93-102. London: Technical Editing and Reproduction Ltd.

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. The Journal of the Acoutical Society of America, 20, 150-159.

Lindsay, P. H. and Norman, D. A. (1977). Human Information Processing: An Introduction. New York: Academic Press.

McCormick, E. J. (1976). Human Factors in Engineering and Design (4th ed.). New York: McGraw-Hill.

McKinley, R. L. (1980). Voice Communication Research and Evaluation System, Technical Report No. AFAMRL-TR-80-25. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory, DTIC No. ADA088100.

McKinley, R. L. (1981). Voice communication research and evaluation system. Aural Communication in Aviation, AGARD-CP-311, 1.1-1.7. London: Technical Editing and Reproduction Ltd.

McKinley, R. L. and Moore, T. J. (1986). Effect of audio bandwidth and bit error rate on PCM, ADPCM, and LPC speech coding algorithm intelligibility. In the AGARD Conference Proceedings Information Management and Decision Making in Advanced Airborne System, AGARD-CP-414. London: Technical Editing and Reproduction Ltd.

Massaro, D. W. (1970). Perceptual auditory images. Journal of Experimental Psychology, 85, 411-417.

Mayer, M. S. (1985). The computerized Diagnostic Rhyme Test as a design tool for armored vehicle intercommunications systems. In the Proceedings of the 1985 IEEE Military Communications Conference (MILCOM '85), 166-170. Boston, Massachusetts: Institute of Electrical and Electronic Engineers.

Miller, G. A. and Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. American Journal of Psychology, 63, 176-185.

Moore, T. J. (1981). Voice communications jamming research. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 2.1-2.6. London: Technical Editing and Reproduction Ltd.

Moore, T. J. and McKinley, R. L. (1986). Research on speech processing for military avionics. In the Proceedings of the Human Factors Society 30th Annual Meeting, 1331-1335. Santa Monica, California: Human Factors Society.

Moore, T. J., McKinley, R. L., and Mortimer, V. D. (1979). Comparative Intelligibility of Standardized Test Materials Processed by the ARC-164 and ARC-34 Radio Systems in the Presence of Simulated Cockpit Noise, Technical Report No. AMRL-TR-79-108. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Moore, T. J., McKinley, R. L., Mortimer, V. D., and Nixon, C. W. (1978). Evaluation of Word Intelligibility of Two Modulator/Demodulator Systems of a Spread Spectrum Communication System in Presence of Simulated Cockpit Noise, Technical Report No. AMRL-TR-78-54. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory.

Moray, N. (1959). Attention in dichotic listening. Quarterly Journal of Experimental Psychology, 11, 56-60.

Moray, N. (1967). Where is attention limited. A survey and a model. Acta Psychologica, 27, 84-92.

Moray, N. (1979). Mental Workload: Its Theory and Measurement. New York: Plenum Press.

Moray, N. (1982). Subjective mental workload. Human Factors, 24, 25-40.

Moray, N. (1988). Focused and divided attention. In K. Boff (Chair), Engineering of Man-Machine Systems: Human Performance for Systems Designers. Symposium conducted at the University of Dayton, Dayton, Ohio.

Moray, N., Johanssen, J., Pew, R. D., Rasmussen, J., Sanders, A. F., and Wickens, D. C. (1979). Report of the experimental psychology group. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press.

Navon, D. and Gopher, D. (1979). On the economy of the human-processing system. Psychological Review, 86, 214-255.

Nixon, C. W. (1984). Voice Communication and Positive Pressure Breathing in Noise, Technical Report No. AFAMRL-TR-84-009. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Nixon, C. W., McKinley, R. L., and Moore, T. J. (1982). Increase in jammed word intelligibility due to training of listeners. Aviation, Space, and Environmental Medicine, 53, 239-244.

Norman, D. A. (1968). Toward a theory of memory and attention. Psychological Review, 75, 522-536.

Norman, D. A. (1976). Memory and Attention: An Introduction to Human Information Processing (2nd ed.). New York: Wiley.

North, R. A., Stackhouse, S. P., and Graffunder, K. (1979). Performance, Physiological, and Oculometer Evaluation of VTOL Landing Displays, Report No. 3171. Langley, Virginia: NASA Langley Research Center.

O'Donnell, R. D. (1979). Contributions of Psychophysiological Techniques to Aircraft Design and Other Operational Problems, AGARD-AG-244. London: Technical Editing and Reproduction Ltd.

O'Donnell, R. D. and Eggemeier, F. T. (1983). Conceptual framework for the development of workload metrics in sustained operations. In Sustained Intensive Air Operations: Physiological and Performance Aspects, AGARD-CP-338, 4.1-4.10. London: Specialised Printing Services Limited.

O'Donnell, R. D. and Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Performance, Vol. II. Cognitive Processes and Performance (Chapter 42). New York: Wiley and Sons.

O'Donnell, R. D. and Wilson, G. (1987). Cortical Evoked Response and Eyeblink Measures in the Workload Evaluation of Alternative Landing System Displays. In A. H. Roscoe (Ed.), The Practical Assessment of Pilot Workload, AGARD-AG-282, 52-55, London: Specialised Printing Services Limited.

Ogden, G. D., Levine, J. M., and Eisner, E. (1979). Measurement of workload by secondary tasks. Human Factors, 21, 529-548.

Ostry, D., Moray, N., and Marks, G. (1976). Attention, practice, and semantic targets. Journal of Experimental Psychology: Human Perception and Performance, 12, 326-336.

Perez, W. A., Masline, P. J., Ramsey, E. G., and Urban, K. E. (1987). Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology, Technical Report No. AAMRL-TR-87-007. Wright-Patterson Air Force Base, Ohio: Armstrong Aerospace Medical Research Laboratory.

Pollack, I. (1958). Speech intelligibility at high noise levels: Effect of short-term exposure. The Journal of the Acoustical Society of America, 30, 282-285.

Porubcansky, C. A. (1985). Speech technology: Present and future applications in the airborne environment. Aviation, Space, and Environmental Medicine, 56, 138-143.

Pratt, R. L. (1981). An automated multiple choice intelligibility testing system. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 20.1-20.3. London: Technical Editing and Reproduction Ltd.

Prohaska, A. M. and Nixon, C. W. (1984). Hearing Protection and Speech Intelligibility of Two Nonstandard In-Flight Communications Headsets, Technical Report No. AFAMRL-TR-84-504. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory, DTIC No. ADA146110.

Reagan, D. M. (1977). Steady-state evoked potentials. Journal of the Optical Society of America, 671, 1475-1483.

Reid, G. B., Shingledecker, C. A., and Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. In the Proceedings of the Human Factors Society 25th Annual Meeting, 522-526. Santa Monica, California: Human Factors Society.

Rolfe, J. M. (1971). The secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.), Measurement of Man at Work, 135-148. London: Taylor and Francis.

Roscoe, A. H. (Ed.). (1978). Assessing pilot workload, AGARD-AG-233. Neuilly Sur Seine, France: Advisory Group for Aerospace Research and Development.

Schmidt-Nielsen, A. (1987). The effect of narrow-band digital processing and bit error rate on the intelligibility of ICAO spelling alphabet words. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-35, 1101-1115.

Schultz, W. C., Newell, F. D., and Whitbeck, R. F. (1970). A study of relationships between aircraft system performance and pilot ratings. In the Proceedings of the Sixth Annual NASA-University Conference on Manual Control, 339-340, Dayton, Ohio: Wright State University.

Shannon, C. E. (1964). The mathematical theory of communication. In C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, 29-129. Urbana, Illinois: The University of Illinois Press (Original work published in 1948, Bell System Technical Journal, July-October, 301-344).

Shannon, C. E. and Weaver, W. (1964). The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois Press (Original work published 1949).

Shiffrin, R. M. (1976). Capacity limitations in information processing, attention and memory. In W. K. Estes (Ed.), Handbook of Learning and Cognitive Processes, Volume 4: Memory Processes, 177-236. Hillsdale, New Jersey: Lawrence Erlbaum.

Shingledecker, C. A. (1983). Behavioral and subjective workload metrics for operational environments. In Sustained Intensive Air Operations: Physiological and Performance Aspects, AGARD-CP-338. Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Shingledecker, C. A. (1984). A Task Battery for Applied Human Performance Assessment Research, Technical Report No. AFAMRL-TR-84-071. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Slowiaczek, L. M. and Nusbaum, H. C. (1985). Effects of speech and pitch contour on the perception of synthetic speech. Human Factors, 27, 701-712.

Spieth, W., Curtis, J. F., and Webster, J. C. (1954). Responding to one of two simultaneous messages. The Journal of the Acoustical Society of America, 18, 418-424.

Squires, K. C., Wickens, C., Squires, N. K., and Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event related potential. Science, 193, 1142-1146.

Steeneken, H. J. M. (1987). Diagnostic information from subjective and objective intelligibility tests. In the Proceedings: ICASSP 87, 1987 International Conference on Acoustics, Speech, and Signal Processing, 5-8. Dallas, Texas: Institute of Electrical and Electronic Engineers.

Steeneken, H. J. M. and Houtgast, T. (1980). A physical method for measuring speech-transmission quality. The Journal of the Acoustical Society of America, 67, 381-326.

Steeneken, H. J. M. and Houtgast, T. (1981). Some applications of the STI-Method in evaluating speech transmission channels. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 18.1-18.6. London: Technical Editing and Reproduction Ltd.

Systems Research Laboratories, Inc. (1987). NEWAMI User's Guide. Dayton, Ohio: Systems Research Laboratories, Inc. (Unpublished).

Treisman, A. M. (1964). Verbal cues, langauge, and meaning in selective attention. The American Journal of Psychology, 77, 206-219.

Treisman, A. M. (1969). Strategies and models of selective attention. Psychological Review, 76, 282-292.

Van Cott, H. P. and Kinkade, R. G. (Eds). (1972). Human Engineering Guide to Equipment Design (rev. ed.). Washington, D.C.: U.S. Government Printing Office.

Voiers, W. D. (1968). The present state of digital vocoding technique. A diagnostic evaluation. IEEE Transactions on Audio and Electronics, AU-16, 275-279.

Voiers, W. D. (1977). Diagnostic evaluation of speech intelligibility. In M. E. Hawley (Ed.), Speech Intelligibility and Speaker Recognition (Benchmark papers in acoustics, Vol. 11, 374-387). Stroudsburg, PA: Dowden Hutchinson & Ross.

von Gierke, H. E. and Nixon, C. E. (1985). Vibration, noise, and communication. In R. L. DeHart (Ed.), Fundamentals of Aerospace Medicine, 250-298. Philadelphia, Pennsylvania: Lea and Febiger.

Weaver, W. (1964). Recent contributions to the mathematical theory of communication. In C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, 1-29. Urbana, Illinois: The University of Illinois Press (Original work published in 1949, Scientific American, July, 1-28).

Webster, J. C. (1965). Speech communications as limited by ambient noise. The Journal of the Acoustical Society of America, 37, 692-699.

Webster, J. C. (1979). Effects of noise on speech. In C. M. Harris (Ed.), Handbook of Noise Control (Chapter 14). New York: McGraw-Hill.

Wheeler, P. D. and Halliday, S. G. (1981). An active noise reduction system for aircrew helmets. In K. E. Money (Ed.), Aural Communication in Aviation, AGARD-CP-311, 22.1-22.8. London: Technical Editing and Reproduction Ltd.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIII. Hillsdale, New Jersey: Erlbaum.

Wickens, C. D. (1984a). Engineering Psychology and Human Performance. Columbus, Ohio: Merrill Publishing.

Wickens, C. D. (1984b). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Wickens, C. D. (1986). The effects of control dynamics on performance. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance. Vol. II. Cognitive Processes and Pereformance (Chapter 39). New York: Wiley.

Wickens, C. D. (1988). Manual Control. In K. Boff (Chair), Engineering for Man-Machine Systems: Human Performance for Systems Designers. Symposium conducted at the University of Dayton, Dayton, Ohio.

Wickens, C. D. and Derrick, W. (1981). Workload measurement and multiple resources. In the Proceedings of IEEE Conference on Cybernetics and Society, 600-603. New York: Institute of Electrical and Electronic Engineers.

Wickens, C. D. and Yeh, Y. Y. (1983). The dissociation of subjective ratings and performance: A multiple resources approach. In the Proceedings of the Human Factors Society 27th Annual Meeting, 244-248. Santa Monica, California: Human Factors Society.

Wierwille, W. W. and Casali, J. G. (1983a). The Sensitivity and Intrusion of Mental Workload Estimation Techniques in Piloting Tasks, Report No. 8309. Blacksburg, Virginia: Virginia Polytechnic Institute and State University, Vehicle Simulation Laboratory, Department of Industrial Engineering and Operations Research.

Wierwille, W. W. and Casali, J. G. (1983b). A validated rating scale for global mental workload measurement applications. In the Proceedings of the Human Factors Society 27th Annual Meeting, 129-133. Santa Monica, California: Human Factors Society.

Wierwille, W. W. and Connor, S. A. (1983). Evaluation of twenty pilot workload measures using a psychomotor task in a moving-base aircraft simulator. Human Factors, 25, 1-16.

Williams, C. E., Forstall, J. R., and Greene, J. W. (1971). The use of in-flight manikin recordings for evaluating the communication effectiveness of flight helmets. IEEE Transactions on Audio and Electroacoustics, AU-19, 97-102.

Williams, C. E., Mosko, J. D., and Greene, J. W. (1976). Evaluating the ability of aircrew personnel to hear speech in their operational environments. Aviation, Space, and Environmental Medicine, 47, 154-158.

Williges, R. C., and Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21, 549-574.

Wolfe, J. D. (1978). Crew Workload Assessment: Development of a Measure of Operator Workload, Technical Report No. AFDL-TR-78-165. Wright-Patterson Air Force Base, Ohio: Air Force Flight Dynamics Laboratory.